SMALL-AREA STATISTICS PAPERS
SERIES GE-41, No. 5

METHODOLOGY AND USE OF SMALL-AREA STATISTICS IN DECISIONMAKING AND 1977 ECONOMIC CENSUSES AND THEIR USE IN PRIVATE AND PUBLIC SECTORS

Papers Presented at the Conference on Small-Area Statistics

American Statistical Association
San Diego, Calif.
August 15, 1978

U.S. Department of Commerce
BUREAU OF THE CENSUS

SMALL-AREA STATISTICS PAPERS
SERIES GE-41, No. 5

METHODOLOGY AND USE OF SMALL-AREA STATISTICS IN DECISIONMAKING AND 1977 ECONOMIC CENSUSES AND THEIR USE IN PRIVATE AND PUBLIC SECTORS

Papers Presented at the Conference on Small-Area Statistics

American Statistical Association
San Diego, Calif.
August 15, 1978

Issued May 1979

U.S. Department of Commerce
Juanita M. Kreps, Secretary
Courtenay M. Slater, Chief Economist

BUREAU OF THE CENSUS
(Vacant), Director

Library of Congress Cataloging in Publication Data

Conference on Small-Area Statistics, San Diego, Calif., 1978.
Methodology and use of small-area statistics in decisionmaking and 1977 economic censuses and their use in private and public sectors.

(Small-area statistics papers : Series GE-41 ; no. 5)
1. United States—Statistical services—Congresses. 2. United States—Census—Congresses. I. American Statistical Association. II. Title. III. Title: 1977 economic censuses and their use in private and public sectors. IV. Title: Small-Area Statistics. IV. Series.
HA37.U55C64 1978a 001.4'22 79-607028

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402. Postage stamps not acceptable; currency submitted at sender's risk. Remittances from foreign countries must be by international money order or by draft on a U.S. bank.
Stock No. 003—024—01871—3

PREFACE

This report contains the papers presented at the conference on Small-Area Statistics in San Diego, California, on August 15, 1978, during two sessions of the annual meeting of the American Statistical Association (ASA), which was held jointly with the Biometric Society and the Institute of Mathematical Statistics.

The first session of the 1978 Conference concerned *Methodology and Use of Small-Area Statistics in Decisionmaking.* John H. Morawetz organized and chaired this session. The speakers were Jacob Silver, Tyler R. Sturdevant, Charles H. Ptacek, Richard S. Conway, Jr., and Manual Cárdenas. Richard C. Taeuber served as the discussant for the first two papers and Jonah Otelsberg served as discussant for the next two papers, there was no discussant for Manual Cardenas' paper.

The second session dealt primarily with the *1977 Economic Censuses and Their Use in the Private and Public Sectors.* Edward J. Spar organized and chaired this session. The speakers were Shirley Kallek, Elias Fokas, Malcolm M. Knapp, and John T. Snow. Evelyn S. Mann and William J. Hawkes, Jr. were discussants for all four papers. (Neither discussant submitted a written paper, therefore, none are presented in this report.)

This report was organized and prepared under the direction of Jacob Silver, Chief, Geography Division, Bureau of the Census. Assisting in the preparation were Ms. Jane Green and Ms. Nancy James.

CONTENTS

METHODOLOGY AND USE OF SMALL-AREA STATISTICS IN DECISIONMAKING

1977 ECONOMIC CENSUSES AND THEIR USE IN PRIVATE AND PUBLIC SECTORS

Methodology and Use of Small-Area Statistics in Decisionmaking

Introduction

John H. Morawetz
McGraw-Hill Informations Systems Company

Let us remember that the profession of "Statistician" had its origin when the need arose in England to have someone collect data for the State, i.e., for the country. Only in the recent past has there been a rapidly growing interest in small-area data. The Small-Area Data Committee within the American Statistical Association (ASA) is, therefore, a recent innovation.

Today, small-area data have become a necessity for the public as well as the private sector of the economy. The Federal, State, and local governments could not administer their various programs without detailed, reliable, and current data. The business community wants to know more about their local markets. While they could survive without such data, they would not be doing as well.

In the spirit of this growing need, we especially welcome the many thoughful papers that will be presented from this podium today.

GBF/DIME Files— A Geographic Tool for Small Area Data

Jacob Silver
Bureau of the Census

INTRODUCTION

I would like to start my presentation with a series of questions which are very common in the public and private sectors of our community today:

1. Where do people live and where do people work?
2. Where are the recipients of aid to dependent children located within the city and suburbs?
3. Where do school age children live in relation to the school they attend?
4. What changes have occurred in the geographical patterns of crime incidents?
5. Is the distribution of home repair loans the same as savings account customers?

In our day-to-day activities, we are finding that organizing local data into meaningful geographic units and analyzing their spatial patterns is becoming more and more an essential requirement in both public and private sector activities and programs. Because of this, there has been an increasing demand—a need in today's "statistical" society—for an effective, computerized geographic referencing, and geographic coding system to assist in providing this type of information; one which is well documented and standarized, but flexible in use. One such system has been developed by the Census Bureau and is known as the GBF/DIME System.

BACKGROUND

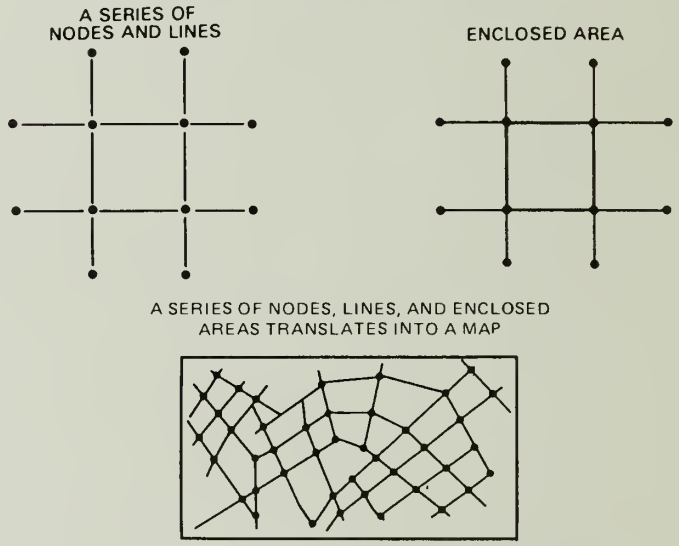
Before I proceed, I should give you a brief background on the development of the GBF/DIME System and its evolvement into a national system. Computerized, geographic coding systems are not a new phenomenon. A number of systems were developed during the early 1960's by transportation and planning agencies. Unfortunately, they were ahead of their time. Most of the files developed during this period were not fully utilized and frequently lacked the necessary financial and technical support. This is no longer the case.

In 1970, the Bureau of the Census conducted the Nineteenth Decennial Census of Population and Housing by a combination of two methods: a mail-out/mail-back system in the larger urban areas of the Nation and a house-to-house enumeration in the remainder of the country. For the urban cores of 145 standard metropolitan statistical areas in which the mail-out/mail-back procedures were used, a method was needed to code individual addresses to specific geographic units for tabulation purposes. With the cooperation of local councils of government and regional and county planning agencies, a geographic referencing system was developed to code approximately 35 million addresses, by computer, to the appropriate geographic areas. The geographic coding file developed at that time was referred to as the Address Coding Guide.

While the Address Coding Guide was sufficiently accurate for geographic coding of the questionnaires, there were certain limitations to the file. First, it had been developed in a format that did not permit a more comprehensive system of editing techniques, using the capabilities of the computer; second, the file did not contain certain features that would permit greater use of the files. For example, it did not contain X-Y location coordinate values.

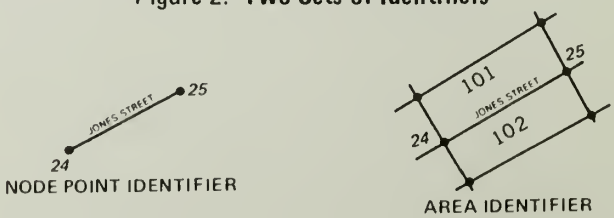
Using graph theory as the conceptual framework, an approach to overcome these problems was developed. This approach combined the address information from the Address Coding Guide files with graph information necessary to describe the urban street network. By considering each street on a map as a series of lines and each intersection of lines as a node point, an entire map sheet can be viewed as a series of interrelated lines, node points, and enclosed areas. (See figure 1.)

Figure 1. Conceptual Framework



This approach is called DIME (Dual Independent Map Encoding). The term DIME refers to the fact that the basic file which represents the map geography is created by defining two independent sets of identifiers for each line segment: (1) the node points at the end of each line (nodes 24 and 25 in figure 2), and (2) the enclosed areas on either side of the line segment (blocks 101 and 102).

Figure 2. Two Sets of Identifiers



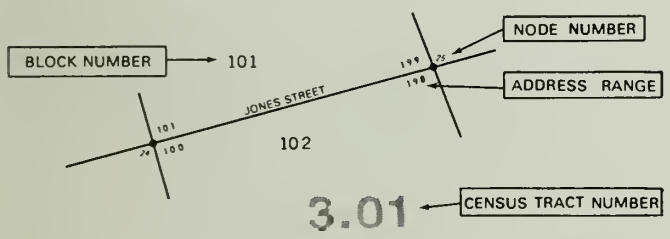
WHAT IS THE GBF/DIME SYSTEM?

A computer file which contains data grouped by geographic area systematically organized is referred to as Geographic Base File (GBF). The Census Bureau's GBF which utilizes the DIME approach is thus referred to as the GBF/DIME-File. The total GBF/DIME System is composed of a computerized geographic reference file, the Census Bureau's Metropolitan Map Series, a series of clerical and computer maintenance programs, and a series of user oriented programs.

The GBF/DIME-File is a computerized version of a map. It contains all features shown on the Census Bureau's Metropolitan Map Series plus block-by-block address ranges, ZIP codes, and X-Y coordinate values (latitude/longitude and State plane) where map features intersect.

Each computer record in the GBF/DIME-File identifies a single segment of a feature between two node points and all of the geographic information related to that segment. A street or non-street feature on the map is divided into a series of segments (or records) as the result of intersecting with other features. In figure 3, Jones Street, between the two node points of 24 and 25, is a segment of that feature and will be a record in the GBF/DIME-File. The address range along the even side is 100 to 198, and along the odd side is 101 to 199. Associated with each side of the segment are the appropriate codes for census block, census tract, place, ZIP, etc. For local purposes, the file could also contain codes for school district, economic neighborhood, or transportation zone.

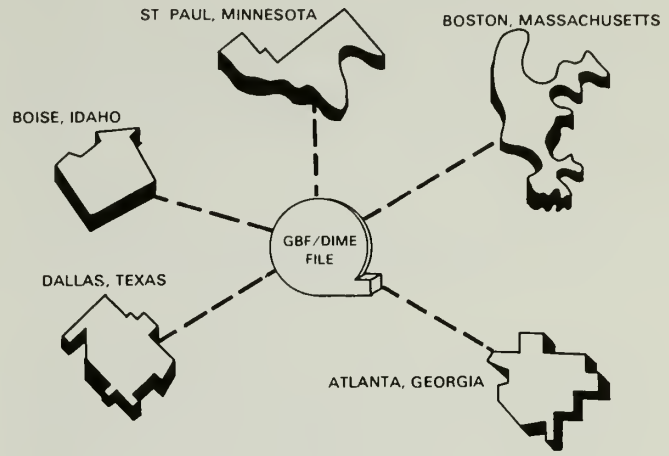
Figure 3. Types of Information



Standardizing the methodology of the GBF/DIME System to systematically define, create, and maintain a current and accurate file and map series is the key ingredient of this nationwide system. Instead of hundreds of independent and largely non-compatible local files, there is one standard system that can be similarly used to develop and maintain a file, and it is comparable whether it is Atlanta, Georgia or Boise, Idaho. (See figure 4.)

However, standardization does not imply that the GBF/DIME System is rigid, inflexible, and identical in format and use in every area of the country. The GBF/DIME System is considered to have two parts: (1) containing certain standard geographically defined elements applicable to all areas, such as street name, address number, block number, and census tract; and (2) containing local geographic elements which will vary from area to area, reflecting local usage and needs such as transportation zone, police beat, school district, or sales district.

Figure 4. Nationwide System—Standardization



THE CUE PROGRAM—A COOPERATIVE EFFORT

The Census Bureau, in preparation for the 1980 Census of Population and Housing, established a cooperative program with local agencies referred to as the CUE Program (Correction, Update, and Extension) of the GBF/DIME System. Under the CUE program, the Census Bureau provides local agencies (mainly councils of government, regional, and county planning agencies), with the clerical procedures, processing methodology, quality control programs, computer programs, and technical assistance necessary to carry out the establishment or the maintenance of the maps and files.

While some agencies are able to clerically determine the changes and additions to street features, political, and statistical boundaries, they may not have available to them the technical personnel or the computer facilities necessary to carry out the computer maintenance phases of the program. Where this situation exists, the Census Bureau carrys out the computer maintenance operation. In addition, as part of the cooperative program, the Bureau inserts the X-Y coordinate values into the GBF/DIME-Files.

The Census Bureau has provided funds through a program of Joint Statistical Agreements (JSA's) to help defray some of the cost incurred by the local agencies in establishing a file where there was none as well as carrying out the correction, update, and extension phases of the program. The Census Bureau was able to fund approved local requests on a 75/25 percentage match ratio, with the local agency matching its 25 percent in either money or services-in-kind. In the last 4 years, the Bureau will have allocated almost 9 million dollars to local cooperative assistance.

We expect to have a GBF/DIME-File for most of the 277 SMSA's in the United States. December 31, 1978, has been established as the latest date for which a corrected/updated GBF/DIME-File can be returned to the Census Bureau for use in preparation for the 1980 Decennial Census. It should be emphasized that the cooperative CUE program with local agencies does not end on that date, but will continue. However,

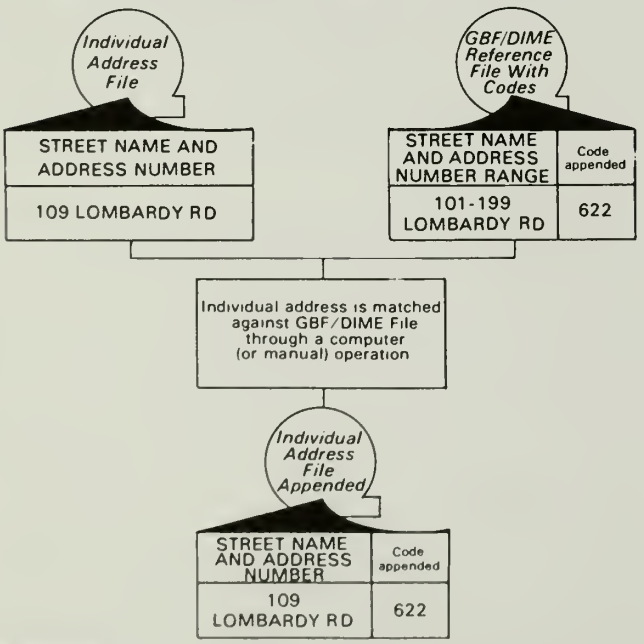
it is essential to the Census Bureau to know far in advance the number, location, and coverage of the areas that will be participating in the GBF/DIME program so that the Bureau can prepare the necessary operational plans for a particular area. For the 1980 Decennial Census, we expect to geographically code, through the use of the computer, 45 million household addresses out of the approximately 86 million addresses to be enumerated.

LOCAL APPLICATION OF THE GBF/DIME SYSTEM

While the GBF/DIME System was originally developed to serve as one of the prime geographic processing and geographic coding resources for the Bureau of the Census, in its decennial census operations, it has become increasingly important to agencies and organizations working at the city, county, and regional levels. It is becoming one of the basic tools used in programs requiring geographic area identification of address-relatable data. The ability to cross-reference (1) street addresses, (2) geographic codes, and (3) X-Y coordinate values provides a flexible geographic "framework" for management and for planning daily operational activities, as well as for research.

The most common application of the GBF/DIME-File is the "address matching" and the "geocoding" of addressed data. In such a case, the GBF/DIME-File is used with an address matching computer program (e.g., ADMATCH AND UNIMATCH*). These programs are designed to accept data records which contain street addresses and to append to the record the geographic unit in which the address is located. Figure 5 illustrates this process.

Figure 5. Address Matching and Geocoding Process



*These programs are available from Customer Services Branch, Data User Services Division, Bureau of the Census, Washington, D.C. 20233.

Address matching is the process of matching records in two files on the basis of street name and address number. In the example, 109 Lombardy Road is matched against the range of addresses in the GBF/DIME-File and is identified as being located in the 101-199 address range of Lombardy Road. Once this match is made, the address (109 Lombardy Road) is assigned all or selected geographic identifiers in the file. (See figure 6.)

Figure 6. Geographic Identifiers

Address	School district	Census tract	Transportation zone	Neighborhood	Etc.
109 Lombardy Rd. . .	12	351.01	622	31	Etc.

This geographic identification of an address is referred to as geocoding. Once a file of individual addresses has been geocoded, the data related to that address can be tabulated along with the data of all the other addresses geocoded to the same geographical unit. For example, the total number of building code violations could be tabulated for each housing district in which they occurred.

Another application of the GBF/DIME-File has been in serving as the digital geographic base for many computer-generated mapping systems. In order to be able to prepare computer maps, the data must be defined by location and the computer must be able to translate this information into relative positions, that is, X-Y coordinates used to define locations.

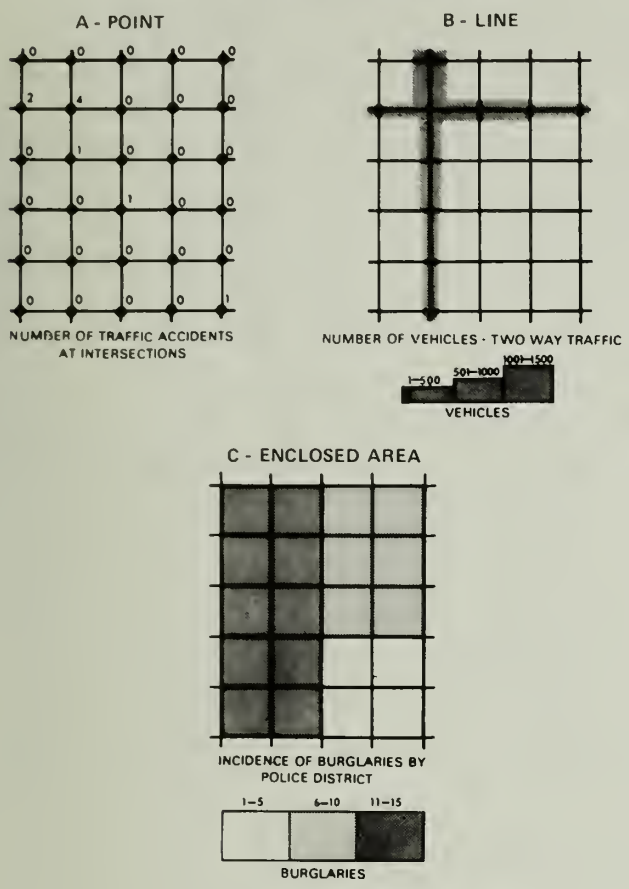
There are actually two coordinate values, latitude/longitude and State plane, available in the GBF/DIME-Files. These values provide the basis for graphically reproducing the map sheets from which it was developed, as well as any data that is related to a node point, a line, or an enclosed area. For example, the incidence of burglaries by block or police district, the amount of two-way traffic along a street segment, or the number of accidents at an intersection, each, can be graphically displayed. (See figure 7.)

The computer-generated map expands the visual communication of tabulated data and provides a spatial analysis of data. Spatial patterns are important in any analysis because they describe the distribution of an activity or incidence, and the patterns illustrated may point to trends in terms of direction, extent, and magnitude between two or more points in time. For example, there are changes that take place in the origin of work trips when there are changes in residential patterns.

The GBF/DIME-File has been successfully used in the areas of education, transportation, emergency services (law enforcement and fire protection), urban planning, health, and welfare, and a wide variety of other public, and private sector programs.

Northeastern Indiana Regional Coordinating Council
Ft. Wayne, Indiana
February 4, 1977

Figure 7. **Computer-Generated Maps**



> . . . its greatest benefit has been as a timesaving device. Information that was hand tabulated and aggregated in the past can now be done with the computer. This has allowed us to use our personnel for other tasks.
>
> Orange County Administrative Offices
> Santa Ana, California
> March 14, 1977

The GBF/DIME file has enabled us to geocode data files previously considered too large to manually geocode. Through an Interactive Spatial Information System, we can access all geocoded variables in a minimal amount of time. This has increased our analytical capabilities because less manpower is required for data collection and aggregation so more emphasis can be applied to socioeconomic analysis.

EXAMPLES OF LOCAL USE

I began this paper by presenting a series of questions which are very common in the public and private sectors of our community. I would like to cite a few examples of the general areas in which the GBF/DIME-Files are being used as one of the inputs to help answer these questions. The following examples are extracted from letters received by the Census Bureau.

Education

School districts are using the system in such programs as:

1. Determining the geographic distribution of school age children by grades, to provide inputs to school facility planning, and to justify capital expenditures.
2. Revising school district boundaries in order to distribute pupil loads more equitably.
3. Optimizing "walk to school" or "bus to school" students, as well as school bus routing.

> Spokane Regional Planning Conference
> Spokane, Washington
> February 24, 1977

Description.—Using the GBF/DIME-File and the ADMATCH computer program, student addresses are geocoded with census tract and block numbers. Block totals are then developed for each sex and grade level. School boundary studies utilize the block totals to determine the effect of shifting boundaries to solve under/over-crowding problems. Transportation planners use the block totals to determine the effect of bus rerouting.

Added capabilities, improvements, etc.—We eliminated manually produced 'pin maps' representing students saving clerical time in every school. Data was more up-to-date in that the pinups were not maintained throughout the school year.

Impact on decisionmaking.—Planners know that accurate information is available and do not hesitate to use it now that it doesn't involve a crash project by each school to update its pin map!

Cost differential.—Saved hours of time for staff of each school.

Transportation

Transportation agencies are using the system in such programs as:

1. Determining distribution of motor vehicles within the community.
2. Determining distribution of residence and employment which is used to study journey-to-work trips.
3. Street network and traffic flow analysis.
4. Carpooling.

> Jackson City Planning Board
> Jackson, Mississippi
> March 10, 1977

The GBF/DIME-File is presently used with our matcher program for the allocation of transportation planning variables for monitoring of the urban area transportation plan. Matching on the address field such variables as auto registration and employment identifiers are allocated to traffic analysis zones which have been previously geocoded, using the block and node system of the GBF/DIME-File. On board ridership surveys have also been matched through the file to sample the origin, destination, and trip length of mass transit ridership. . . .

The GBF/DIME-File, in conjunction with the matcher program, permits a more efficient use of staff time and reduces staff effort. For example, it took two to three persons about 2 weeks, by hand, to match vehicle registrations (approximately 75,000 records). Using the GBF/DIME-File and the matcher program, the effort was reduced to about one-half day and one person.

Emergency Services

Law enforcement and fire protection agencies are using the system in such programs as:

1. Determining geographic patterns of crime and/or fire incidents, in order to optimize the design of police beats, fire stations, and manpower allocation.
2. Computer Aided Dispatching (CAD) systems for police, fire, and ambulance services.
3. Hazardous location recognition.

Metropolitan Area Planning Agency
Omaha, Nebraska
January 18, 1977

Applications of the GBF/DIME file to solve local problems have slowly begun to be implemented. Current applications now being used are. . . .

(a) Creation of an on-line display screen for interactive address and jurisdiction referencing for the Omaha Police Division.

(b) The computer plotting of Omaha Fire Department fire calls during the years 1964, 1971, 1972, and 1973. These maps will be used as part of the comprehensive fire station master plan currently being prepared by the Omaha City Planning Department and the Omaha Fire Division.

Urban Planning

Regional, county, and city planning agencies are using the system in such programs as:

1. Housing condition surveys.
2. New housing starts (building permits) to determine high growth areas.
3. Shifts in neighborhood social-economic characteristics.

Springfield Planning Department
Springfield, Massachusetts
March 8, 1977

The following describes one of its uses (the GBF/DIME System) in the real estate data base sector of the city's information system.

The primary use of the GBF/DIME File is to standardize the geographic bases used within the various city departments and agencies. The planning department maintains an on-line integrated real estate parcel file incorporating the records of five municipal departments. This file is coded to both census and political geography, using the census tract and block for most analytical programs. . . .

The use of the geocoded real estate file has proved to be a valued asset for a number of city applications. Federal programs, for example, often request a description of the housing inventory in various target neighborhoods. The planning department maintains a file of the geography numbers for the various community development target areas and program boundaries. Consequently, one can request a multitude of programs analyzing the area with continuously updated information or request a list of mailing labels with property owners names and addresses. The ability to easily and quickly retrieve data by neighborhoods, enables planners and decisionmakers to have more time to study the information rather than collecting it.

The use of geocoded data has greatly improved the city's ability to provide information and services. Previously, data was provided by address or owners' name, making it sometimes practically impossible to collect and analyze data within the time frames of various projects. The use of geocoded records for notifying property owners is one place where direct savings are easily found. Zone changes and historical commission hearings often require the notification of owners in larger areas and sometimes up to 4,000 properties. In the past, a clerical pool was formed from city hall offices to copy and type envelopes for notifying property owners, taking days to accomplish a job that the computer now does in a couple of hours.

Health and Welfare

Health and welfare services are using the system in such programs as:

1. Distribution of the sick and the aged in relation to health facilities.
2. Optimizing caseworker loads.
3. Determining where new services are needed.
4. Health research.

Orange County Administrative Office
Santa Ana, California
March 14, 1977

All these geocoded data files (aggregated to census tract) are utilized by the Program Planning division of the county administrative office for indepth socioeconomic analysis. The variables displayed by means of choropleth maps, tables, and graphs in a series by State of the county reports.

These reports indicate areas where:

(1) future analysis is required,
(2) shifts in services are required,
(3) new services are needed,
(4) future growth will occur, and
(5) potential socioeconomic problems will develop.

We have used our GBF/DIME-File for ADMATCH processing of numerous local user files. The diversity of files geocoded are as follows:

(1) Community Referral Information System file–

The geocoded file is used to generate annual reports indicating the distribution of clients by category of referral, by age, and by type of contact. . . .

(2) County of Orange Health Department files–

(a) dog license
(b) acute communicable disease
(c) child health
(d) pulmonary disease
(e) maternal health

(3) County of Orange Mental Health Department files–

(a) alcoholism
(b) drug abuse
(c) mental health

Other Uses

Delaware Valley Regional-Planning Commission
Philadelphia, Pennsylvania
February 17, 1977

New Federal regulations require mortgage lending institutions to provide census tract codes for all properties mortgaged. This created a need on the part of the bankers for a mechanism which would geocode property and addresses to tract. The GBF/DIME-File for the Delaware Valley provided this mechanism.

(The file was processed and the results published in a 96 page book which contains a street and address range index.)

The Delaware Valley Regional Planning Commission, in conjunction with the mortgage bankers association, produced the index. The index allows manual geocoding of any address in the nine county area. Sales of the index have paid for the entire cost of production.

SUMMARY

While the GBF/DIME System was originally developed by the Census Bureau as a geographical tool to serve as its major geographic processing and geocoding resource, the usefulness of this tool has also been recognized at all levels of the public as well as private sectors.

Substantial savings are being achieved through the local use of this system with activities which require the knowledge of the spatial location of address-relatable data. More important, the massive amount of local data–spatially identified–can be made more meaningful, more understandable, and more communicative to those in decisionmaking positions–people like the mayor, the councilman, the public health director, the president of a bank, or the chairman of the neighborhood civic committee.

Using Available Resources to Generate Small-Area Data

Tyler R. Sturdevant
Bureau of the Census

INTRODUCTION

Timely economic data are produced usually for only major geographic areas. Small-area statistics are available infrequently and then only on a delayed basis.

In this paper, the lack of current small area information is explored and some of the needs are noted, with particular emphasis on retail trade. Major underlying causes of the situation are then examined and methods explored which would help to alleviate the dilemma. These involve using large area estimates obtained from current sample surveys in conjunction with available relevant small area information. Various techniques are discussed, including the use of synthetic estimates and regression equations. Illustrations and examples in applying the methodology to current retail sales estimates are given. Finally, the possible sources of small area information for retail sales are reviewed and the advantages and limitations which might be encountered in using this methodology to generate small area current estimates of retail sales are discussed.

USES AND AVAILABILITY OF SMALL-AREA DATA

While statistics at national levels are adequate and timely, there are a number of applications for which there is no substitute for small geographic area data.

Some of the needs and lack of availability of small-area data, with emphasis on unemployment and housing, were described recently by Maria Gonzalez and Christine Hoza.[1] Joseph W. Duncan outlined the demands for regional data and pointed out difficulties in supplying them.[2]

The person considering advertising on TV would like to know the size of the retail market reached by a station. The firm which is expanding its outlets may wish to know the approximate local sales for a particular kind of business. A national distributor would be interested in the potential marketing channels for its product.

Users of small area economic data discover that data from economic censuses are available for small areas, but only once each 5 years. In the past, there had been a 2- to 3-year delay from the end of the data year until the census reports are available. For some purposes, 2- to 6-year old data may be adequate, but for dynamic trades this may be quite unsatisfactory. Consider the eating and drinking trade which is normally a high turnover industry. Dramatic changes have taken place during the past decade with rapid expansion of the fast food outlets. Yet in mid-1978, the latest detailed Census Bureau information on this industry is from the 1972 Census of Retail Trade. Is that current enough for decisions to be made in 1978? The obvious answer is that, for some purposes, the information is out of date. Unfortunately, data from the 1977 census will not be available until 1979.

What alternative information is available? Weekly retail trade sales, released the following week, are timely, but represent only major retail kinds of business at the national level. Monthly retail trade advance estimates, released 10 days following the close of the month are similar to the weekly retail survey in scope and coverage, utilizing the same panel of respondents. More detailed estimates of retail sales are available 45 days after the close of the month, and revised a month later based upon additional observations. These provide national information for some detailed kinds of business, lesser detail for geographic regions, divisions, for selected large States, standard metropolitan statistical areas (SMSA's), and cities. Monthly department store sales are available for selected SMSA's, cities, central business districts, and miscellaneous areas, because all known department stores are canvassed monthly.

County Business Patterns, published by the Census Bureau, generally in the second year following the data year, provides county and large city aggregates by detailed kinds of business, including such items as number of establishments, employment, and payroll totals for those firms with employees subject to Social Security taxes.

The Internal Revenue Service publishes statistics of income from individuals, business, and corporate returns, which can be useful. In particular, "small-area data from individual tax returns" is very helpful. Number of returns, number of exemptions, and amounts of income, are classified by size of gross income and are presented for States, counties, and selected SMSA's.

In addition to sources of information directly related to income or sales, there are small area data that may be highly correlated with certain kinds of retail sales. One such example is population estimates made by the Census Bureau. Such estimates are made by States; usually by October of the reference year and; for counties and SMSA's, a year later. Estimates for selected cities are available on a somewhat delayed basis. On the assumption that people tend to make most retail purchases in the areas where they live, good estimates of small area population should be highly correlated with retail sales. This assumption, of course, may not hold true in areas frequented by tourists of characterized by seasonal changes in the resident population.

Other sources of small area data, such as State sales tax information, retail credit association data, and list of licensed firms, might be useful in some application but would be much less readily available and would require a great deal of effort in compilation and interpretation.

---

[1] Gonzalez, Maria Elena and Christine Hoza, "Small Area Estimation with Application to Unemployment and Housing Estimates," *Journal of the American Statistical Association*, Vol. 73, 7-15.

[2] Duncan, Joseph W., "The Demand for Regional and Local Area Statistics: Issues Concerning the National Response," *Statistical Reporter*, Number 78-4, January 1978.

CAUSES FOR DATA INADEQUACIES

While there is need for current small area data, the more current estimates are made at national or large area levels only. Economic censuses are taken once every 5 years, but results have not become available until 2 or 3 years after the data year. What are the underlying reasons for this situation?

Current estimates are usually obtained from sample surveys, and it is well known that for comparable estimation precision, it takes nearly the same size of sample for a small area as for a large one. Therefore, the development of reliable small area estimates from current surveys would require a greatly increased sample size compared to one designed to produce only large area estimates. The barriers to such action are obvious—substantial respondent burden, increased resources of money, personnel, and computer processing time. In fact, Federal guidelines on paperwork reduction specify that surveys conducted less frequently, than annually, may not be designed for the express purpose of producing subnational estimates. This does not mean, of course, that such estimates cannot be derived as byproducts. Samples allocated to produce national estimates for detailed kinds of business at specified levels of precision will yield estimates of comparable precision for broader business classifications for sufficiently large subnational areas.

The quinquennial economic censuses obtain data through mailed questionnaires or use administrative data from the respective universe. The collection process takes about 6 months and is a massive operation. Once collected, the form is screened for completeness, with contact made to the respondent where necessary. Other processing includes keying, editing, and imputation where needed. Supplemental administrative records are available after the close of the data year (October). These records are edited and merged with reported data to provide the complete files for tabulation and disclosure analysis. The huge mass of records, the large number of tables, and data items require considerable time to process and check, which explains the minimum period of 1 year for summary totals to be available and the additional 1 to 2 years to complete the detailed tables.

Similar delays are experienced in processing income tax summaries by IRS and Social security employer tax information by the Census Bureau.

METHODS OF ESTIMATING CURRENT SMALL-AREA DATA

As suggested by the title of this paper, current small area data can be generated from current large area estimates using, in addition, the most recent small area data for which a relationship has been established. The concept is not new. Gonzalez and Hoza[3] reviewed some of the earlier applications by Hansen, Hurwitz, and Madow in 1953;[4] Lillian Madow in 1956;[5] and by Ralph Woodruff in 1969.[6] Most of the more recent developments have been in the demographic fields.

One method of utilizing auxiliary information along with current large area estimates is through the process of synthetic estimates ". . .by assuming that for the statistic of interest the mean value in the large area applies to each subarea directly."[7] Gonzalez also describes a more refined method of making this assumption for subgroups of the population, making sure that subgroups are uniquely defined, nonoverlapping, and exhaustive.

For example, if one is interested in obtaining estimates of total retail trade sales for 1977 by States, a simple estimate could be made as follows:

$X_{2.}$ = total 1972 retail sales in the United States.
X_{2j} = total 1972 retail sales in State j.

then:

$\frac{X_{2j}}{X} = P_{2j}$ = Proportion of 1972 retail sales in the United States represented by State j.

If, in 1977 a sample survey indicates—

$X_{7.}$ = estimated total 1977 retail sales in the United States.

Then, a simple synthetic estimate is:

$X_{7j} = P_{2j} X_{7.}$ estimated 1977 retail sales in State j.

This method assumes that the proportion of national retail sales is static between 1972 and 1977.

To refine the estimate, we may try to obtain a better estimate of P_{7j} than P_{2j}. One method is to observe trends of P_{ij} between 1967 and 1972 for each State, thus projecting the increasing or decreasing trends. Before the projected P_{7j} values are used, the sum should be forced to equal unity. An alternative approach is to use regional or divisional estimates, where available, on the assumption that a localized area is more homogeneous in economic activity than are broader areas.

Rather than assume a State's share of national retail sales will change from 1972 to 1977 at the same rate as from 1967 to 1972, one can make use of additional information, such as population estimates and retail trade employer payroll information. When one begins to specify dependent and independent variables, models using multiple regression equations become appropriate. The first step is to identify the available variables and to construct models for a census year in order to generate appropriate coefficients and to evaluate the validity of the models. using information from the 1960 Census of Population and Housing, Gonzalez and Waksberg constructed estimates of vacancies for subareas in 1970 and were able to estimate the error rates for the subareas, using the root mean square as a measure.[8]

[3] ibid.

[4] Hansen, Morris H., William N. Hurwitz, and William G. Madow (1953) *Sample Survey Methods and Theory*, Vol. 1, John Wiley and Sons.

[5] Madow, Lillian (1956), "U.S. Television Households by Region, State and County—March 1956," Advertising Research Foundation, New York.

[6] Woodruff, Ralph S. (1966), "Use of a Regression Technique to Produce Area Breakdowns of the monthly National Estimates of Retail Trade," *Journal of the American Statistical Association*, Vol. 61, pp. 496-504.

[7] Gonzalez, Maria Elena (1973), "Use and Evaluation of Synthetic Estimates," Proceedings of the Social Statistics Section of the American Statistical Association, pp. 33-36.

[8] Gonzalez, Maria Elena and Joseph Waksberg (1973), "Estimation of the Error of Sythetic Estimates," unpublished paper presented at the first meeting of the International Association of Survey Statisticians, Vienna, Austria.

For building models for retail trade subarea estimates, results from the 1972 Census of Retail Trade may be utilized in conjunction with current estimates for that year and auxiliary information aforementioned.

Initial search for available auxiliary information is still in progress. For illustrative purposes, tables are appended showing various information items for 13 major States and the coefficients of correlation with retail trade sales in those States for 1972. In addition, examples of estimation methods for generating small area data are given, using San Diego county retail sales estimates for 1972.

To give an indication of the relationship of per capita sales for a subarea of a larger region, three charts are presented. In figure 1, you may see the similarity in the trends of per capita retail sales of San Diego county as a percentage of California for three major classifications of nondurable kinds of business. Furthermore, when you consider the lower per capita income of San Diego county in comparison to the entire State, the trends are at plausible levels. Figure 2, gives a similar illustration for three major classifications of durable goods for kinds of business. The building boom in San Diego county, which peaked in 1972, may offer some rationale for building material per capita sales ratio behaving differently between 1967 and 1972 than the other classifications. Figure 3, shows the ratios for three dissimilar industries was added to inject a note of caution in oversimplifying per capita sales relationships. In trying to determine a plausible explanation for the erratic behavior of apparel goods store sales between 1967 and 1972, it was discovered that errors in classification of stores, geographically as well as reporting errors for kind of business, contributed to the apparent plunge in per capita sales for this category for San Diego county. It is interesting to note that editing procedures for the 1977 Census of Retail Trade will include, among others, per capita sales trend changes which would tend to prevent such large classification errors from escaping detection in the 1977 census processing for States and SMSA's.

FACTORS TO BE CONSIDERED IN GENERATING SMALL-AREA DATA

Whether the Census Bureau actually produces current estimates for small areas will depend upon a number of policy decisions. Is there a demonstrated need for such estimates? Is

Figure 1. Percentage of Retail Per Capita Sales of Three Major Classifications for Nondurable Goods

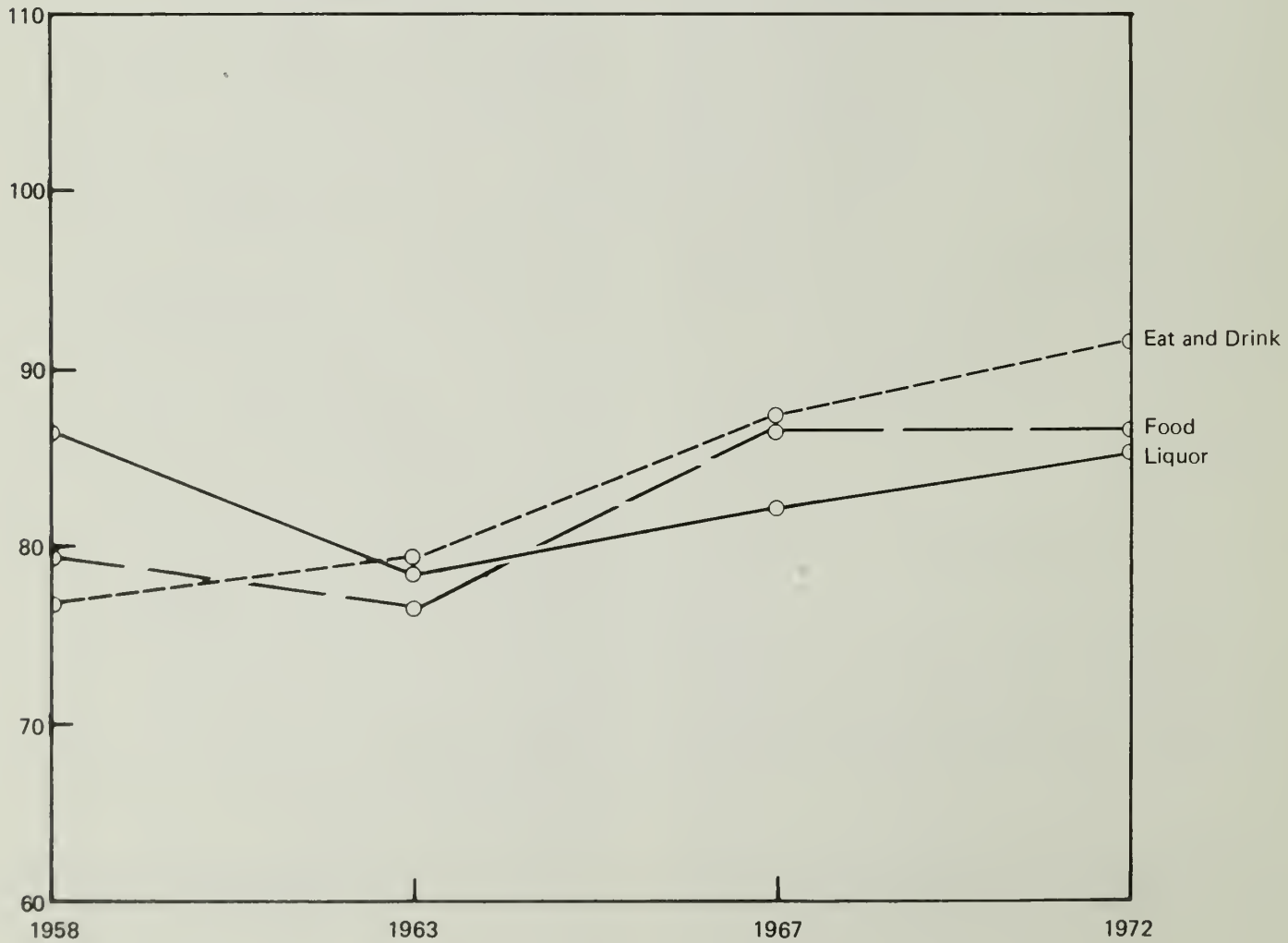
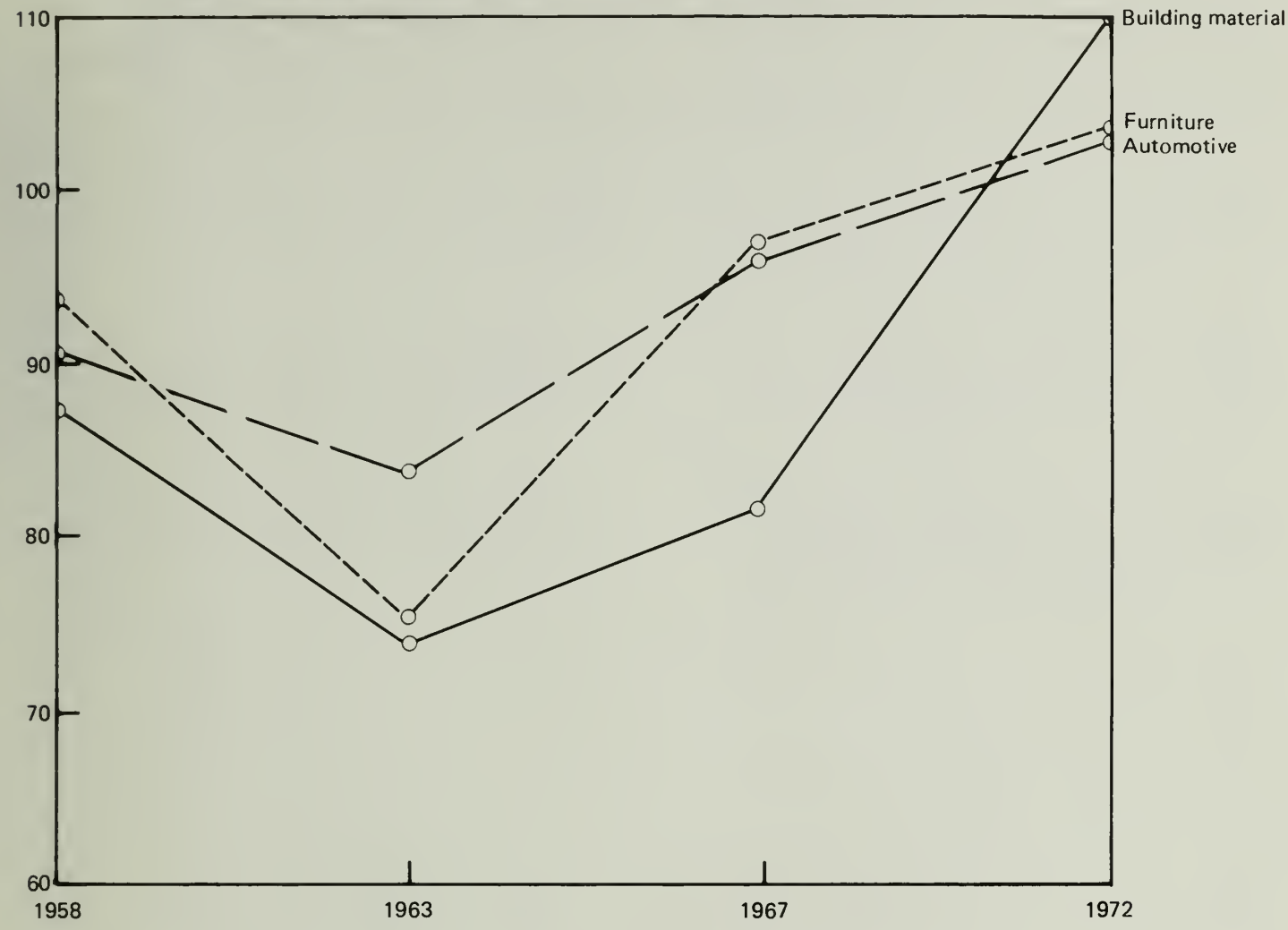


Figure 2. **Percentage of Retail Per Capita Sales of Three Major Classifications for Durable Goods**



demonstration of past reliability of such estimates sufficient justification for release as a Census Bureau estimate? Can resources be found to develop sound estimates?

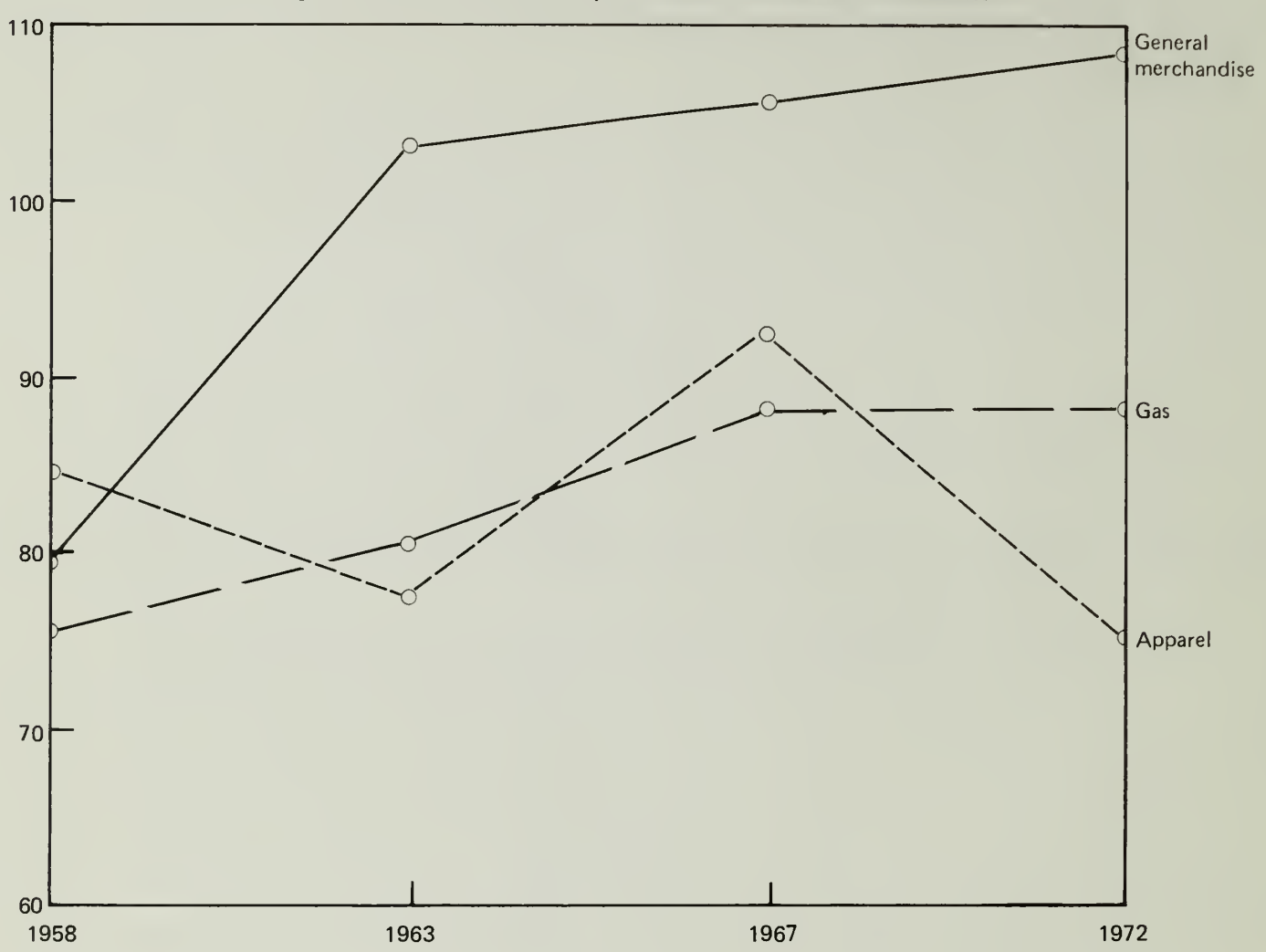
With increasing demands on economy, in government, it seems indefensible to waste information by not using it. As long as useful information can be produced economically and with demonstrated ranges of confidence, the full utilization of available information is highly appealing—particularly to the economists and marketing trade. For some statisticians, however, the idea of the Census Bureau producing estimates with the measurement of bias dependent upon the periodic validation of the estimation model, is offensive.

One of the limiting factors in generating small area data on a monthly basis is that most auxiliary information is on an annual basis. For that reason, variation in seasonality of retail trade sales would be a complicating factor. This may be less of a problem if the large area monthly base estimate is for a geographic area with characteristics similar to the small area for which data are being generated.

Form of presentation is also a factor. Guidelines for publication standards were developed by statisticians at the Census Bureau and published as a separate part of a Journal of the American Statistical association (JASA) volume.[9] These standards call for users to be aware of the lack of reliability of data, e.g., to the extent that underlying model assumptions may no longer be valid. More basic would be the determination of small areas for which data would be estimated. Smaller area data are more desirable, but estimates would be less reliable.

[9] Gonzalez, Maria Elena; Jack L. Ogus; Gary Shapiro; and Benjamin J. Tepping (1975), "Standards for Discussion and Presentation of Errors in Survey and Census Data," *Journal of the American Statistical Association*, Vol. 70, Part II.

Figure 3. **Ratio of Retail Per Capita Sales for Three Dissimilar Industries**



SUMMARY

While there is pressure to relieve respondent burden of reporting, there are also increased demands for more frequent and timely small area data.

The Bureau of the Census is exploring the availability of auxiliary data which could supplement current large area estimates. Research is beginning on developing appropriate estimation models, using results from the 1972 and 1977 economic censuses to measure reliability of estimates. Methods of measuring estimation errors will be patterned after techniques developed in demographic areas of the Bureau, i.e., using mean squares.

Practical cognizance must be taken concerning user need, respondent burden, utilization of resources, maintenance of quality, and standards.

The ultimate decision will be a balanced consideration of responsibility, responsiveness, and resourcefulness.

REFERENCES

1. Duncan, Joseph W., "The Demand for Regional and Local Area Statistics: Issues Concerning the National Response," Statistical Reporter, Number 78-4, January 1978.

2. Gonzalez, Maria Elena (1973), "Use and Evaluation of Synthetic Estimates," Proceedings of the Social Statistics Section of the American Statistical Association, pp. 33-36.

3. Gonzalez, Maria Elena and Christine Hoza, "Small Area Estimation with Application to Unemployment and Housing Estimates," *Journal of the American Statistical Association*, Vol. 73, pp. 7-15.

4. Gonzalez, Maria Elena, Jack L. Ogus, Gary Shaprio, and Benjamin J. Tepping (1975), "Standards for Discussion and Presentation of Errors in Survey and Census Data," *Journal of the American Statistical Association*, Vol. 70, part II.

5. Gonzalez, Maria Elena and Joseph Waksberg (1973), "Estimation of the Error of Synthetic Estimates," unpublished paper presented at the first meeting of the International Association of Survey Statisticians, Vienna, Austria.

6. Hansen, Morris H., William N. Hurwitz, and William G. Madow (1953), Sample Survey Methods and Theory, Vol. 1, John Wiley and Sons.

7. Madow, Lillian (1956), "U.S. Television Households by Region, State, and County—March 1956," Advertising Research Foundation, New York.

8. Woodruff, Ralph S. (1966), "Use of a Regression Technique to Produce Area Breakdowns of the Monthly National Estimates of Retail Trade," *Journal of the American Statistical Association*, Vol. 61, pp. 496-504.

Table 1. Comparison of Retail Trade Sales and Related Data for the United States and 13 Selected States: 1967 and 1972

State	Census of retail trade sales					1972 current survey total sales	1972 census of population estimates	Internal revenue service			County Business Patterns	
		1972						1972			1972	
	1967 total (million dollars)	Total (million dollars)	Department stores (million dollars)	Food stores (million dollars)	Eating and drinking (million dollars)	(million dollars)	(thousands)	Adjusted gross income (million dollars)	Salaries and wages (million dollars)	Total exemptions (thousands)	Number of employees (thousands)	First quarter payroll (million dollars)
	C1	C2	C3	C4	C5	S1	P1	I1	I2	I3	B1	B2
United States...	310,214	470,806	51,084	100,719	36,868	448,379	208,234	749,228	616,706	210,587	11,642	14,094
California.....	33,498	49,633	5,972	10,652	4,588	46,979	20,416	79,660	65,946	20,905	1,199	1,652
Florida.....	10,280	19,761	2,239	4,011	1,575	18,067	7,390	26,534	19,642	7,468	495	582
Illinois.....	19,252	26,597	3,046	5,207	2,255	26,186	11,216	46,489	38,592	11,562	692	901
Indiana.....	8,329	11,870	1,332	2,401	904	11,604	5,282	18,517	15,502	5,285	296	338
Massachusetts.....	9,167	13,386	1,499	2,905	1,227	12,299	5,790	23,439	19,382	6,200	383	447
Michigan.....	14,114	20,938	2,494	4,710	1,610	19,792	9,010	34,590	29,737	8,824	474	603
Missouri.....	7,561	10,706	1,184	2,125	773	10,935	4,749	15,760	12,777	4,587	280	333
New Jersey.....	11,362	16,934	2,078	3,948	1,424	16,399	7,329	30,778	25,901	7,500	400	526
New York.....	29,091	39,975	4,790	9,526	3,813	37,889	18,367	79,454	65,026	19,121	994	1,373
North Carolina.....	6,648	10,972	853	2,307	601	10,023	5,240	16,222	13,618	5,480	258	297
Ohio.....	16,295	23,272	3,404	5,073	1,906	22,702	10,733	39,154	33,371	11,109	600	711
Pennsylvania.....	17,497	25,627	2,894	5,665	1,903	24,173	11,884	42,067	35,452	11,627	638	747
Texas.....	16,449	26,480	2,671	5,533	1,782	24,720	11,618	36,765	29,566	11,599	682	764

Description of Codes as Shown Above

Code	Data item	Source	Date available
C1	1967 total retail sales	Census Bureau: 1967 Census of Business	October 1969
C2	1972 total retail sales	Census Bureau: 1972 Census of Retail Trade	January 1974
C3	1972 department store sales	Census Bureau: 1972 Census of Retail Trade	January 1974
C4	1972 food store sales	Census Bureau: 1972 Census of Retail Trade	January 1974
C5	1972 eating and drinking places sales	Census Bureau: 1972 Census of Retail Trade	January 1974
S1	Total retail sales	Census Bureau: Monthly Survey of Retail Trade	February 1973
P1	July 1, 1972 population count	Census Bureau: Current Population Reports	September 1972
I1	1972 adjusted gross income	Internal Revenue Service: Small-Area Data	May 1977
I2	1972 salaries and wages	Internal Revenue Service: Small-Area Data	May 1977
I3	1972 number of exemptions	Internal Revenue Service: Small-Area Data	May 1977
B1	1972 number of employees, mid-March pay period, retail trade	Census Bureau: County Business Patterns	October 1973
B2	1972 taxable payrolls, January-March, retail trade	Census Bureau: County Business Patterns	October 1973

Note: All retail trade data are based upon 1967 SIC codes.

Table 2. Correlation and Regression Coefficients Between Selected Data for 13 Selected States

Variables		Coefficient of correlation	d	B	Variables		Coefficient of correlation	d	B
x_1	x_2				x_1	x_2			
C2	C1	.98919	1,445	1.390	I1	P1	.98400	-3,988	4.194
C2	C3	.98344	2,118	7.797	I2	P1	.98229	-3,283	3.465
C2	C4	.99225	1,219	4.376	I3	P1	.99872	-145	1.032
C5	C4	.98489	-259	0.433					
C2	S1	.99893	-118	1.057	C2	B1	.99570	-485	40.927
C2	P1	.99152	-227	2.318	C2	B2	.99333	2,519	28.405
C4	P1	.99439	-305	0.527	S1	B2	.99444	2,486	26.891

Table 3. **Comparison of 1972 Census of Retail Trade Sales With Estimates Generated From Related Data**

State	1972 census Y	Synthetic Y'_1	Regression Y'_2	Regression Y'_3	Regression Y'_4	Regression Y'_5	2 stage regression Y'_6	Regression Y'_7
California.....	49,633	48,418	48,007	49,539	47,832	47,097	46,966	48,585
Florida.....	19,761	14,859	15,734	18,979	18,771	16,903	16,929	19,774
Illinois.....	26,597	27,827	28,205	27,561	24,005	25,772	25,751	27,835
Indiana.....	11,870	12,039	13,022	12,147	11,726	12,017	12,067	11,629
Massachusetts.....	13,386	13,250	14,187	12,882	13,931	13,194	13,235	15,190
Michigan.....	20,938	20,400	21,063	20,802	21,830	20,728	20,732	18,914
Missouri.....	10,706	10,929	11,955	11,440	10,518	10,781	10,837	10,975
New Jersey.....	16,934	16,422	17,238	17,216	18,495	16,762	16,784	15,886
New York.....	39,975	42,048	41,881	39,931	42,905	42,348	42,240	40,196
North Carolina.....	10,972	9,609	10,686	10,476	11,314	11,919	11,962	10,074
Ohio.....	23,272	23,553	24,095	23,878	23,418	24,652	24,635	24,071
Pennsylvania.....	25,627	25,290	25,766	25,433	26,009	27,320	27,291	25,626
Texas.....	26,480	23,775	24,309	26,011	25,431	26,704	26,679	27,427

Description of Variables as Shown Above

Variables	Item
$Y = C_2$	1972 Census of Retail Trade sales
$Y'_1 = \frac{C_1(\text{State})}{C_1(\text{U.S.})} \cdot S_1$	Synthetic estimate based upon State proportion of retail sales in 1967 sales in 1967
$Y'_2 = 1.390\ C_1 + 1{,}445$	Regression of 1967 retail sales on 1972 Census of Retail Trade sales
$Y'_3 = 1.057\ S_1 - 118$	Regression of 1972 current survey sales on 1972 Census of Retail Trade sales
$Y'_4 = 4.376\ C_4 + 1{,}219$	Regression of food store sales on 1972 Census of Retail Trade sales
$Y'_5 = 2.318\ P_1 - 227$	Regression of population estimate July 1, 1972 on 1972 Census of Retail Trade sales
$Y'_6 = 4.376\ C'_4 + 1{,}219$	Two stage regression of population estimate July 1, 1972 on 1972 Census of Retail Trade sales
where	
$C'_4 = 0.527\ P_1 - 305$	
$Y'_7 = 40.927\ B_1 - 485$	Regression of number of retail employees, mid-March, on 1972 Census of Retail Trade sales

Table 4. Comparison of 1967 and 1972 Retail Trade Sales and Selected Data Between the United States, San Diego County, and the State of California

Description	Census of retail trade sales				Current survey estimates sales				Population estimates July 1
	Total (million dollars)	Department stores (million dollars)	Food stores (million dollars)	Eating and drinking (million dollars)	Total (million dollars)	Department stores (million dollars)	Food stores (million dollars)	Eating and drinking (million dollars)	(thousands)
1967									
United States.....	310,214	32,344	70,251	23,843	313,500	27,703	72,137	24,887	197,500
San Diego County.....	1,881	319	414	165	(NA)	256	(NA)	(NA)	1,198
California.....	33,498	3,936	7,647	2,333	32,605	(NA)	(NA)	(NA)	19,176
San Diego, percent of--									
California.....	5.62	8.10	5.41	7.07	-	-	-	-	6.25
United States.....	0.606	0.986	0.589	0.692	-	0.924	-	-	0.607
1972									
United States.....	470,806	51,084	100,719	36,868	448,379	46,302	95,020	33,891	208,200
San Diego County.....	3,310	460	645	295	(NA)	437	(NA)	(NA)	1,443
California.....	49,633	5,972	10,652	4,588	46,979	(NA)	(NA)	(NA)	20,411
San Diego, percent of--									
California.....	6.67	7.70	6.06	6.43	-	-	-	-	7.07
United States.....	0.703	0.900	0.640	0.800	-	0.944	-	-	0.693

- Represents zero. NA Not available.

Source: U.S. Department of Commerce, Bureau of the Census, the 1967 and 1972 Census of Retail Trade, current survey estimate from the Monthly Retail Trade Report, and population estimate from Current Population Reports.

EXAMPLES OF GENERATING SMALL-AREA DATA: 1972 RETAIL SALES ESTIMATES FOR SAN DIEGO COUNTY

Total Retail Sales

$Y'_{SD} = 3{,}310$ (1972 Census of Retail Trade)

Method 1: 1967 proportion San Diego County of California from 1967 census applied to 1972 California total retail sales from current survey estimates:

$$Y'_{SD1} = \frac{1{,}881}{33{,}498} \times 46{,}979 = 2{,}638$$

Method 2: Modifying method 1 for change in proportion of State's population.

$$Y'_{SD2} = 2{,}638 \times \frac{1{,}443}{20{,}411} \times \frac{19{,}176}{1{,}198} = 2{,}985$$

Method 3: Ratio of population increase 1972 over 1967 applied to 1967 census data, adjusted for increase in State per capita total retail sales.

$$Y'_{SD3} = \frac{1{,}443}{1{,}198} \times \frac{1{,}881}{} \times \frac{46{,}979}{20{,}411} \times \frac{19{,}176}{32{,}605} = 3{,}067$$

Method 4: Ratio of department store sales increase, current survey estimates, applied to 1967 census total retail sales.

$$Y'_{SD4} = \frac{437}{256} \times 1{,}881 = 3{,}211$$

Department Store Sales

$Y''_{SD4} = 460$ (1972 Census of retail Trade)

Method 1: Use current survey estimate

$$Y''_{SD1} = 437$$

Method 2: Ratio of census to current 1967 department store sales applied to 1972 current survey estimate

$$Y''_{SD2} = \frac{319}{156} \times 437 = 545$$

Method 3: 1967 San Diego County per capita department store sales applied to 1972 San Diego County population estimate. Adjusted for changes in U.S. per capita department store sales.

$$Y''_{SD3} = \frac{319}{1{,}198} \times 1{,}443 \times \frac{46{,}302}{208{,}200} \times \frac{197{,}500}{27{,}703} = 609$$

Food Store Sales

$Y'''_{SD} = 645$ (1972 Census of Retail Trade)

Method: 1967 San Diego County per capita food store sales applied to 1972 San Diego County population estimate. Adjusted for changes in U.S. per capita food store sales.

$$Y'''_{SD1} = \frac{414}{1{,}198} \times 1{,}443 \times \frac{95{,}020}{208{,}200} \times \frac{197{,}500}{72{,}137} = 623$$

Eating and Drinking Sales

$Y''''_{SD} = 295$ (1972 Census of Retail Trade)

Method: 1967 San Diego County per capita eating and drinking sales applied to 1972 San Diego County population estimate. Adjusted for changes in U.S. per capita eating and drinking sales.

$$Y''''_{SD1} = \frac{165}{1{,}198} \times 1{,}443 \times \frac{33{,}391}{208{,}200} \times \frac{197{,}500}{24{,}887} = 257$$

Discussant

Richard C. Taeuber
Department of Health, Education, and Welfare

I'll begin my brief remarks with the summary comment that I find both papers informative and well written. Both get at preliminary aspects of the use of quantitative information in subnational and local decisionmaking, an important trend, for it is preferable that decisions be made on some basis other than intuition or guess. The two papers, however, address entirely different aspects of small-area data use. Silver's paper on the DIME system goes to the very smallest area: acquiring microdata and allocating it to very small geographic areas. The Sturdevant paper describes what to do when you do not have direct small-area data. The question of what is a "small area" is treated differently—for Silver, it means census tracts and for Sturdevant, it means counties. If you use the Sturdevant approach, you really have no need for the DIME system because, if you're talking of San Diego County, most people could assign specific addresses to the county without use of an address reference system.

For those wanting subcounty data, the DIME system is a very valuable tool; however, there are questions which should be posed. One is the extent of information on the currency of the DIME files—especially the currency of the files to be used as part of the 1980 census. Keeping such a file current may require a massive amount of information. There is also the problem of having the file available and current as of a specific data date. Does the wide use and constancy of the DIME files permit standardized software? To what extent does the Bureau provide and support the use of software, either directly or by case study applications? I found the use of Atlanta as one of the illustrative cities cited intriguing because I had occasion, in the middle of this decade, to work with the 1970 Atlanta DIME file. Displaying individual tracts or ED's on a scope proved to be an easy way of showing a mistake in one or more segments, especially if you do not have closure. Even when you have closure, you can compare the shape with a map for indicated mistakes.

The illustrations in the paper and the slides make me wonder how sensitive is the system to misspellings or use of abbreviations—road (or RD) or the names of streets which can have multiple spellings? The two different spellings might be very valid, or they might be a mistake.

In using the DIME system with the 1980 census, there are a couple of things that cause me some concern. One is the problem of nonbounded urbanized areas—those that have room to grow and change boundaries. What percentage of those areas is lost when you close out file updates in 1978? I can see closing the DIME files in 1978 because of the preparation time required to use them in the data acquisition portion of the census, but when you come subsequently to analysis, to use the DIME file with new data linked to census geography and/or data of the 80's, you will want a DIME file which is current as of the data date; i.e., April 1, 1980, rather than December 31, 1978, or some date prior to that. Servicing the double uses in conjunction with the same data set is an interesting challenge.

The Sturdevant paper points out that national data may be fine, but programs—be they Federal Government, local government, or commercial marketing programs—are implemented in very small areas. National data may provide overall trends, but if you want to market a specific product or implement a specific program, you really want to get to smaller areas: States, sub-States or even subcities. Even in formula grants, to some extent, we are trying to go more and more beyond just allocating to the States—revenue sharing goes down to any political jurisdiction, although I still haven't figured out how a town's mayor can meaningfully spend $250 when $50 of it has to be spent on the paperwork necessary to get the money in the first place.

A national program needs consistent small-area data. You can not have each jurisdiction making its own estimate or using its own synthetic estimates. You must have consistent, small-area data, and that means census data, not just the economic data illustrated in the paper which are available on a 5-year program, but also the population data which are now moving onto a 5-year program as well. Obviously, a single, small area can use local data and get a better estimate than most national programs. But again, if you want consistency, which most federal and other governmental units would want, as would most major corporations, you must use data systems and estimation systems which are not dependent on something peculiar to an individual area.

Sturdevant comments that synthetic estimates and other approaches are needed to cut down lag time. The realities of data delays mean that a decision made today requires a forecast or projection which uses data which may be out of date. One advantage to the approach, at least in the paper, is that it gets some of the background, underlying assumptions in these estimates out into the open. Contrast this with programs which used the cost-of-living estimates where early 1960 shopping patterns were used as weights well into the 1970's.

The use of regression instead of straight ratio estimates may not be as advantageous as claimed. The switch changes the assumption from constancy of the ratios to constancy of the regression coefficiency; it does not eliminate assumptions that the patterns are consistent. In modifying techniques like this, care must be taken that you do not merely mask your problems and assumptions by adding multiple computational layers between the answer and the implicit assumptions in your system.

One sentence in the Sturdevant paper, not mentioned in the oral presentation, needs comment. The paper makes the statement, "For some statisticians, however, the idea of the Census Bureau producing estimates with the measurements of bias depending upon the periodic validation of the estimation model is offensive." I am bothered, because this implies, at least on the part of some, a willingness to assume constancy of an estimating model. Any estimating model has to be challenged periodically. You can not just set it up and then continue to use

it without periodic reexamination—not necessarily changing it, but at least periodically challenging its assumptions and structure.

As a final comment, Sturdevant mentions in his paper that the Bureau is exploring the availability of auxiliary data, which could supplement current large area estimates and enhance the use of synthetic estimation. I would urge also, as a strong personal bias, that they explore user support for materials to show people how to use this technique or how to set up similar systems on their own. For example, people in San Diego might want to use synthetic estimates based on a State data series which is more current. The support here should not be pitched toward the major Governments, nor toward the major national corporations, because they have the resources to explore on their own. The need is to help local users create synthetic estimates for their own small-area needs. This help might be case studies, illustrations where to find data materials, and showing people how to use the technique. Presumably the focus would be on data from the Bureau of the Census, but they should use State-based situations.

Estimating Local Trading Area Potential for the Electrical Contractor Market—Using the Product Potential Method

Charles H. Ptacek
General Electric Company

The market collapse of the large project construction market during the 1974-75 recessionary period identified the need for developing accurate annual estimates for local trading area contractor market potential. Few sources of information are available for estimating trading area market potentials. The primary objective of this study was to develop a reliable method for estimating annual electrical distributor expenditures for the electrical contractor business segment, Electrical Work (SIC 1731). An empirical model composed of four construction modules (building, nonbuilding, residential maintenance and repair, industrial/commercial maintenance and repair) was developed which adequately represented the relationship between the total dollar value of all types of local construction and the dollar value associated with electrical products installed by electrical contractors. The predictive validity associated with the empirical model was assessed by comparing 19 local trading area estimates with local area criterion information. The model's accuracy was judged to be acceptable.

Accurate estimates of local market potentials are needed to effectively and efficiently allocate dollars and resources to the various elements of a supplier's marketing mix. (See references 1 and 2.) Product, pricing, distribution, promotion, labor force, and service elements will vary across local trading areas. Not only do these elements vary between trading areas, they will also shift and change over-time within a trading area. The volatile nature of local trading area spending requires distributors and manufacturers to continually update their estimates of local area market potentials and to identify specific areas for corrective marketing action. Shifts in market potential are caused by a number of factors including or reflecting the following:

- Changes in competitive environments—contractors/distributors
- Changes in buying power—small, medium, and large contractors
- Changes in buying patterns—materials needed
- Changes in local area economic conditions
- New governmental regulations—local and national

The market collapse of the large project construction during the 1974-75 recessionary period identified the need for accurately estimating local trading area contractor market potentials. Shifts in small, medium, and large electrical contractor buying patterns during this period produced shifts in building material demands. Electrical distributors and manufacturers supplying contractor materials found themselves with declining sales and in some cases declining market shares because of their heavy emphasis on project construction where the major market collapse occurred.

RESEARCH OBJECTIVES AND OVERALL STUDY APPROACH

The results presented in this report are extracted from a larger study which provided methods for estimating local trading area market potentials for the electrical contractor, industrial, and commercial market segments.

The primary objective for this phase of study was to develop a reliable method for estimating annual electrical distributor expenditures for the electrical contractor business segment (SIC 1731). Estimates were to be accurate within ±20 percent at the 90 percent confidence level. Local trading areas were defined by counties conforming to the territory boundary lines set forth by management.

An extensive secondary information search was conducted to determine what methods and sources of information were currently being used to estimate contractor market potentials. Over 100 informed sources were contacted, including all levels of government, professional associations, and electrical trade press. Two different methods were developed and validated for estimating local contractor expenditures. This report will present information relating to one method.

CRITERION VARIABLE—MARKET ASSESSMENT STUDIES

Market assessment studies were conducted in seven large trading areas for the purpose of establishing criterion estimates of annual dollar expenditures for the electrical contractor. The trading areas represented a wide geographical pattern and major market cities with a mix of different types of businesses. In addition, these studies provided information relating to local competitive environments and buying practices for the various business segments.

Over 750 indepth telephone interviews were conducted with key buying influences in small, medium, and large electrical contractor establishments. Lists of local area contractor establishments were obtained from two sources—Dunn and Bradstreet and Rickard Publishing Company. A stratified random sample of small, medium, and large electrical contractors were sequentially sampled in each trading area to assure optimum efficiency, accuracy, and representation.

PRODUCT POTENTIAL APPROACH

One method which can be used to estimate local electrical contractor market potential is to estimate the local demand for electrical products. (See reference 3.) The basic assumption associated with this approach is that empirical models can be developed which adequately represent the relationships between

the total dollar value of all types of local construction and the dollar value associated with electrical products installed by electrical contractors. The relationship and, hence, empirical models will depend on a number of factors which include: Type of construction (e.g., building and nonbuilding), building structure types (e.g., single family and hospital), product category, and trading area.

The methodology begins with a construction statistics data base[1] which measures local construction activity at the contract award or comparable stage. Monetary value, floor space, and in some cases dwelling unit information is provided for 271 different structure categories. For example, a years construction activity for one trading area has been conveniently collapsed into 28 structure categories as shown in table 1.

Table 1. Projects Causing Demand Product Category I

Type of construction	Value (million dollars)	Floor space (million sq. ft.)	Total dwelling units
Building			
Single family	156,932	6,146	4,678
2 to 4 families	11,082	463	664
Garden apartments	10,953	643	644
High rise apartments	3,000	75	115
Hotels and motels	3,156	92	(NA)
Dormitories	1,060	12	(NA)
Industrial buildings	49,089	657	(NA)
Office buildings	99,450	947	(NA)
Warehouses	8,209	808	(NA)
Garages and service stations	1,277	50	(NA)
Stores and restaurants	19,608	507	(NA)
Religious buildings	9,806	251	(NA)
Educational	81,012	1,521	(NA)
Hospital	65,570	293	(NA)
Other nonresidential	81,216	1,285	(NA)
Nonbuilding			
Telephone and telegraph	1,225	(NA)	(NA)
Railroads	19,922	(NA)	(NA)
Electric utilities	1,401	(NA)	(NA)
Gas	905	(NA)	(NA)
Petroleum pipelines	37	(NA)	(NA)
Water supply	5,130	(NA)	(NA)
Sewer supplies	62,056	(NA)	(NA)
Highways	55,395	(NA)	(NA)
Military	459	(NA)	(NA)
Conservation and development	2,996	(NA)	(NA)
Other nonbuildings	12,090	(NA)	(NA)
Residential maintenance and repair			
Additions and alterations	336,257	(NA)	(NA)
Maintenance and repairs	354,288	(NA)	(NA)

NA Not available.

[1] Dodge construction statistics data base is produced and developed by McGraw-Hill Information Systems Company.

STRUCTURAL MODEL

As noted in figure 1 the total dollar value of all types of construction can be classified into four modules: building, nonbuilding, residential maintenance and repair, and commercial/industrial maintenance and repair. The basis for this classification rests with the type of empirical model employed with each type of construction. Product types have also been classified into four categories. The basis for this classification relates to the different timelags between the onset of construction and the need for each type of product. For example, product category I has the shortest timelag and includes products which are usually bought and installed early in construction (e.g., wire, cable, and conduit). The timelags will also depend on the size and type of building under construction. Thus, for each product category, a size x type timelag matrix must be developed and applied to the original construction data base. The time dimension is important if one wishes to develop product forecasts or establish market potentials for a fixed or extended period of time.

EMPIRICAL MODEL—BUILDING MODULE

Each of the four types of construction require a different model for estimating market potential. The building module employees an empirically based "factor use" model. Dollar per square foot factors for electrical products have been developed for over 40 different types of building structures. For purposes of this report only 15 building structures will be differentiated. These electrical product use factors were developed via primary and secondary information. Local material cost adjustments and regional adjustments are developed for each trading area via secondary information and applied to each square foot factor. Locality adjusted electrical product use factors across all categories range from \$.18 per sq. ft. to \$1.48 per sq. ft. The overall average is \$.55 with a standard deviation of \$.26.

The building structure square foot totals, as shown in table 1, are multiplied by their locality adjusted product use factors, the sum across the 15 building categories (units in thousands of dollars) is the total electrical contractor market potential for the building module. The building module "factor use" model is summarized as follows:

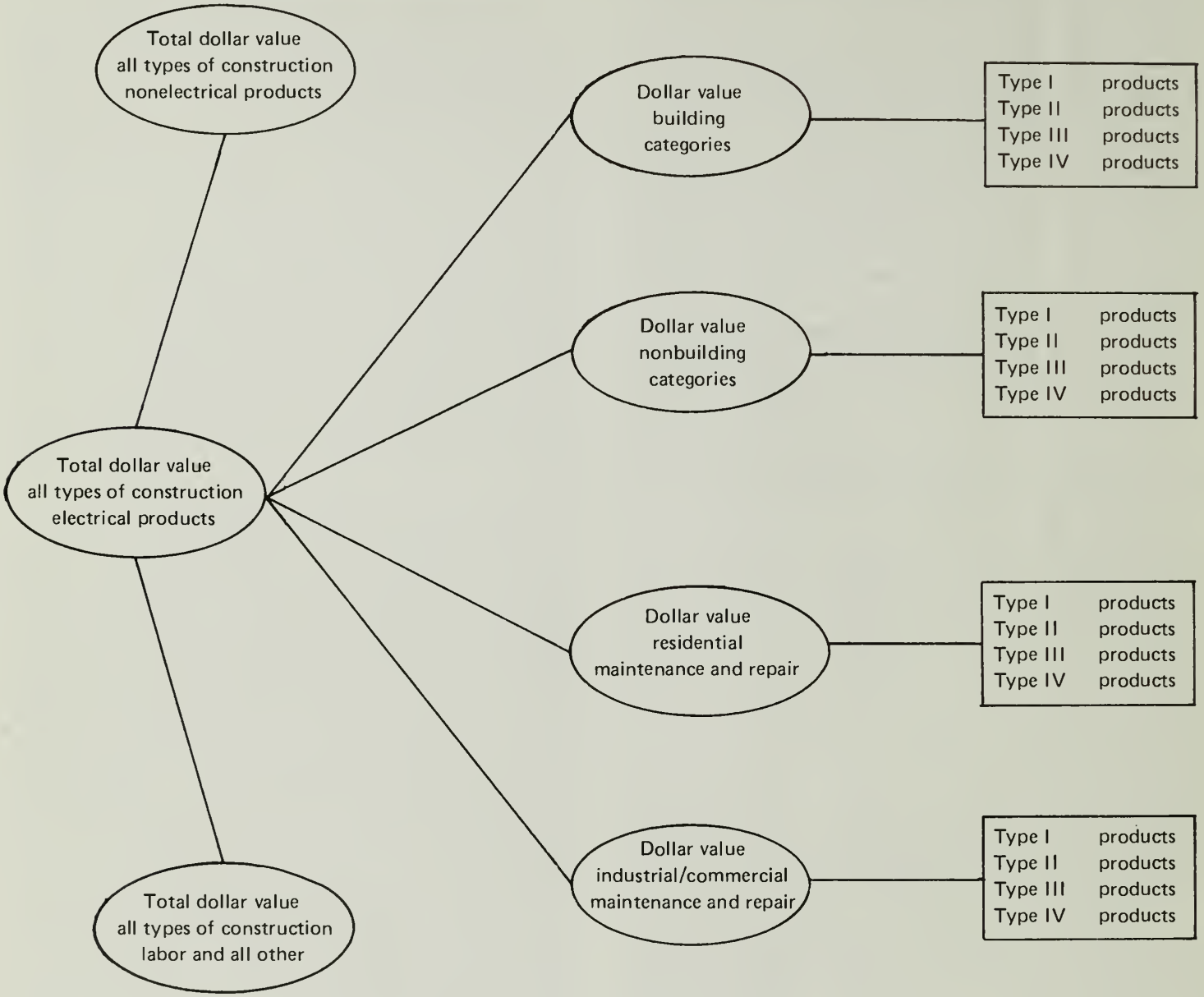
$$\begin{matrix}\text{Local trading area}\\ \text{Building module}\\ \text{Market potential}\end{matrix} = \sum_{i=1}^{15} (PF_i LA_i)\ (SF_i)$$

Where:

- i = Structure type
- PF = Product use factor associated with 15 building categories
- SF = Sq. ft. associated with 15 building structures
- LA = Locality adjustment factors associated with 15 building categories

The building module serves as the core module for the empirical model used to estimate the total local trading area electrical contractor market potential.

Figure 1. Structural Model for Estimating Local Trading Area Contractor Market Potential



EMPIRICAL MODEL—INDUSTRIAL/COMMERCIAL MAINTENANCE AND REPAIR MODULE

Unfortunately, no construction data base exists for the industrial/commercial maintenance and repair module, however, the seven market assessment studies provided information relating to this module. A coefficient was analytically developed relating the dollar potential of the industrial/commercial maintenance and repair to the dollar potential for the building module. The coefficient is applied directly to the total building dollar potential to derive an estimate of electrical contractor dollar potential for the industrial/commercial maintenance and repair module.

EMPIRICAL MODEL—NONBUILDING AND RESIDENTIAL MAINTENANCE AND REPAIR MODULES

Estimation of the nonbuilding and residential maintenance and repair modules require an iterative procedure. The estimate for the nonbuilding module is made by first summing the dollars that apply to these categories and dividing the nonbuilding dollars sum by the dollar sum of all 28 categories. This procedure produces a local trading area nonbuilding "product use" factor. The first approximation for the nonbuilding dollar potential is made by applying the electrical product "use factors" to the residential maintenance and repair categories and adding this dollar sum to the total electrical contractor market

potential for the building module and then applying the nonbuilding product "use factor."

The next step is to develop a more appropriate local dollar potential estimate for the residential maintenance and repair module. This is done by incorporating information obtained in the seven trading area assessment studies with previously calculated information and developing a regression coefficient which will predict local trading area residential maintenance and repair "product use" factors. The analytically derived local residential maintenance and repair estimates result in local market potential estimates for the residential maintenance and repair module.

The residential maintenance and repair estimates are then replaced with the analytically derived estimates and a more accurate estimate for the nonbuilding module is calculated using the same procedure presented earlier. This iterative procedure is employed for each trading area—the only generalized source of information related to the regression coefficient used to develop residential maintenance and repair product "use factors."

The final model developed for estimating local trading area electrical contractor market potential includes four modules and can be summarized as follows:

Local Trading Area Electrical Contractor Market Potential

Building module

$$\left[\sum_{i=1}^{15} (PF_i\ LA_i)\ (SF_i)\right]$$

Industrial/commercial maintenance and repair module

$$+\left[\text{IC Coeff. (BM)}\right]$$

Nonbuilding module*

$$+\left[\left(\frac{\sum_{i=1}^{11} \$NB_i}{\sum_{i=1}^{28} \$AC_i}\right)(BM + \$RM)\right]$$

Residential maintenance and repair module

$$+\left(\sum_{i=1}^{2} (PF_i\ \$RM_i)\right)$$

*Iterative produced described in context.

Where

- i = Structure type
- PF = Produce use factor
- SF = Square feet associated with 15 building structures
- LA = Locality adjustment figures
- $NB = Dollars for nonbuilding construction
- $AC = Dollars for all 28 structure types
- $RM = Dollars for residential maintenance and repair construction
- IC = Regression coefficient for the industrial/commercial maintenance and repair module
- BM = Market potential estimate building module

EVALUATING THE EMPIRICAL MODEL'S ACCURACY FOR ESTIMATING LOCAL TRADING AREA ELECTRICAL CONTRACTOR MARKET POTENTIAL

As noted earlier, market assessment studies were conducted in seven large trading areas for the purpose of establishing criterion estimates of annual dollar expenditures for the electrical contractor market. The product potential method was initially evaluated by comparing the empirical model's local trading area market potential estimates with the trading area criterion information. Figure 2 illustrates the relationship between the model's estimates and the criterion estimates of annual dollar expenditures for the electrical contractor market. The criterion dollar volume estimates are proprietary, however, in figure 2, market potential estimates based on the empirical model were judged to be within ± 10 percent of the local criterion variable for all seven test markets. Although absolute dollar volume estimates can not be diverged it is interesting to note that the dollar volume is approximately equal for all four modules of the empirical model.

There are, however, several limitations associated with this evaluation. First, it should be noted that the trading area criterion variable is really nothing more than an estimate of a true population parameter. This is noted by the confidence bands surrounding the criterion estimates in figure 2. Thus, both the product potential estimates and criterion estimates could be consistently inaccurate. Second, the product potential estimates and criterion estimates are not mutually independent, see maintenance and repair modules. Third, the evaluation was made by comparing estimates from seven of the largest trading areas in the United States. It seems possible that the empirical model could produce inaccurate market potential estimates for smaller trading areas.

As noted earlier, an alternative approach was also developed and validated. This approach is based on conducting annual mail surveys with electrical contractors and electrical distributors. The method relies heavily on developing accurate territorial index parameters (i.e., dollar per employee indexes). As illustrated in figure 3, method B also produced accurate estimates within ± 15 percent of the criterion variable. The importance of this finding relates to the fact that the product potential method and method B produced similar market potential estimates for the seven test markets. Thus, substantiating the initial evaluation with an independent methodology. This finding also allows for further evaluations to be made in smaller trading areas. The assumption is that product potential method and method B will produce market potential estimates within ± 10 percent for the smaller trading areas. Figure 4 indicates that the product potential method and method B produce consistent electrical contractor market potential estimates for medium and small trading areas. This type of

Figure 2. **Product Potential and Method B Test Market Validation Using Criterion Confidence Market Estimate as Base**

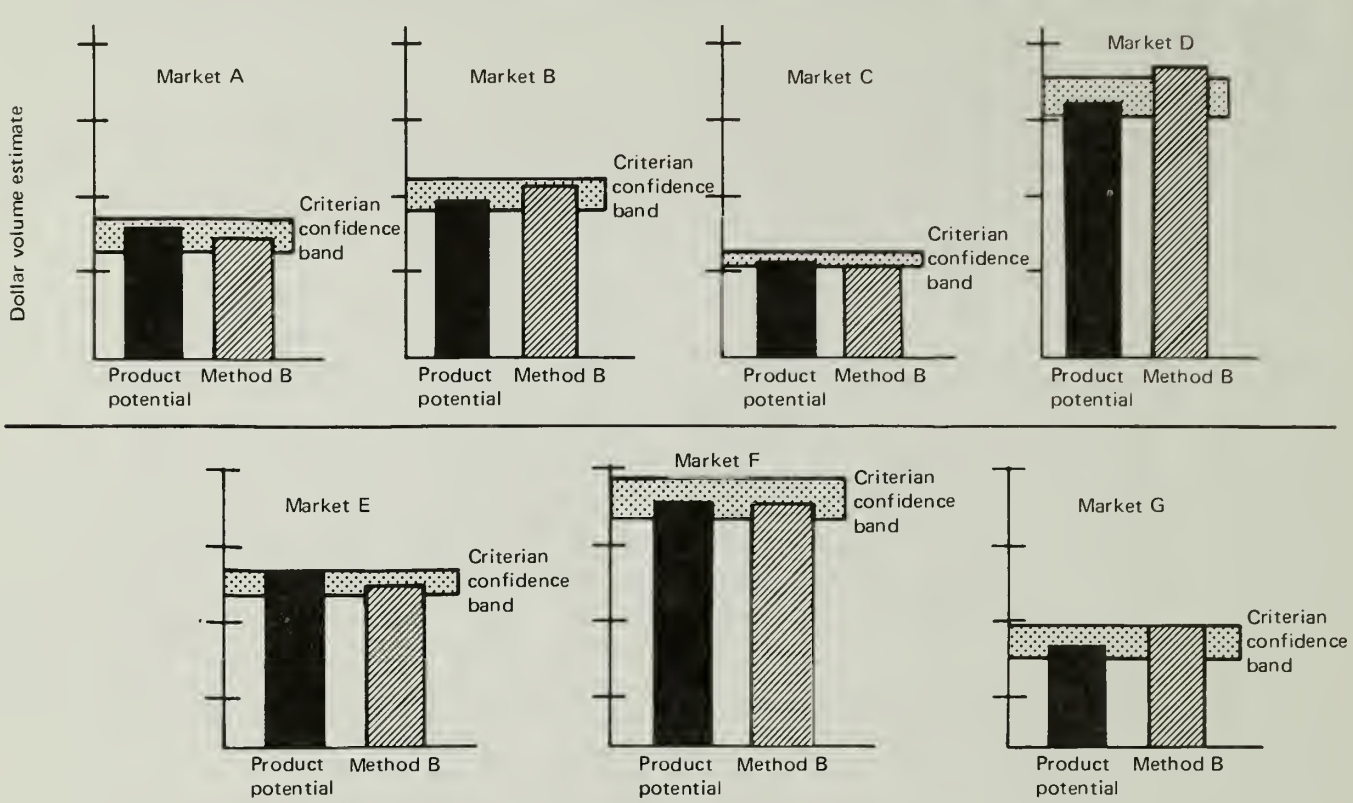


Figure 3. **Dodge and Method B Consistency Validation—Small Markets Six Markets Electrical Contractor Market Potential**

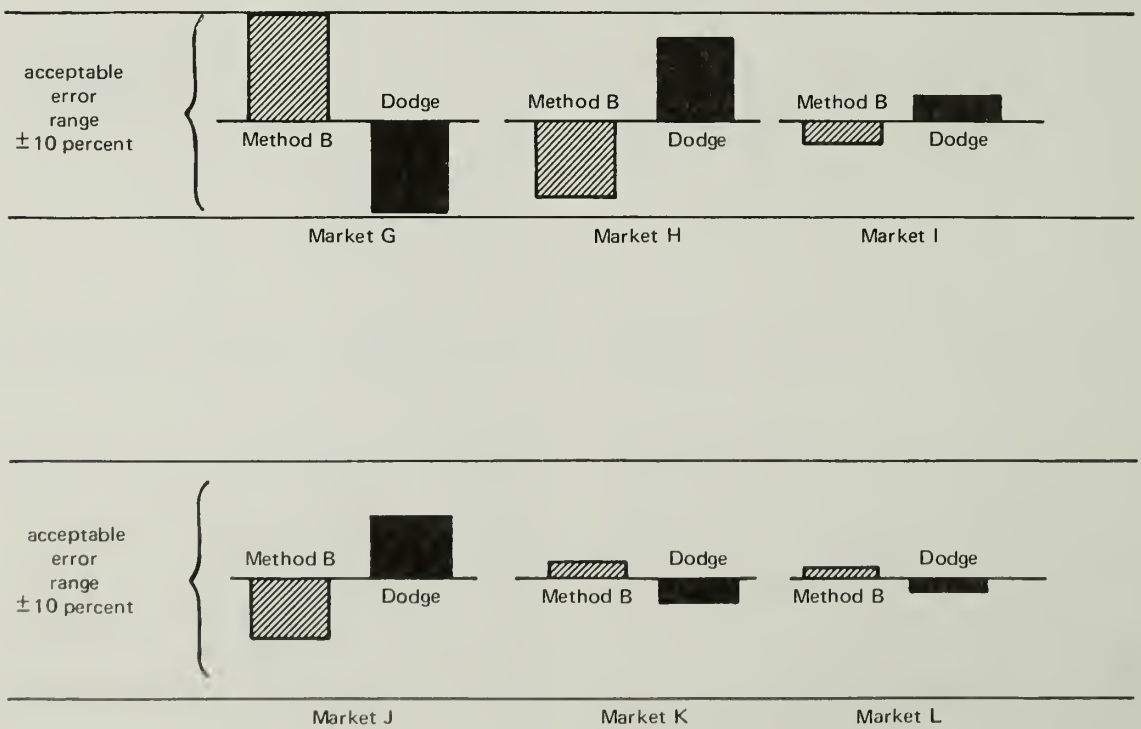
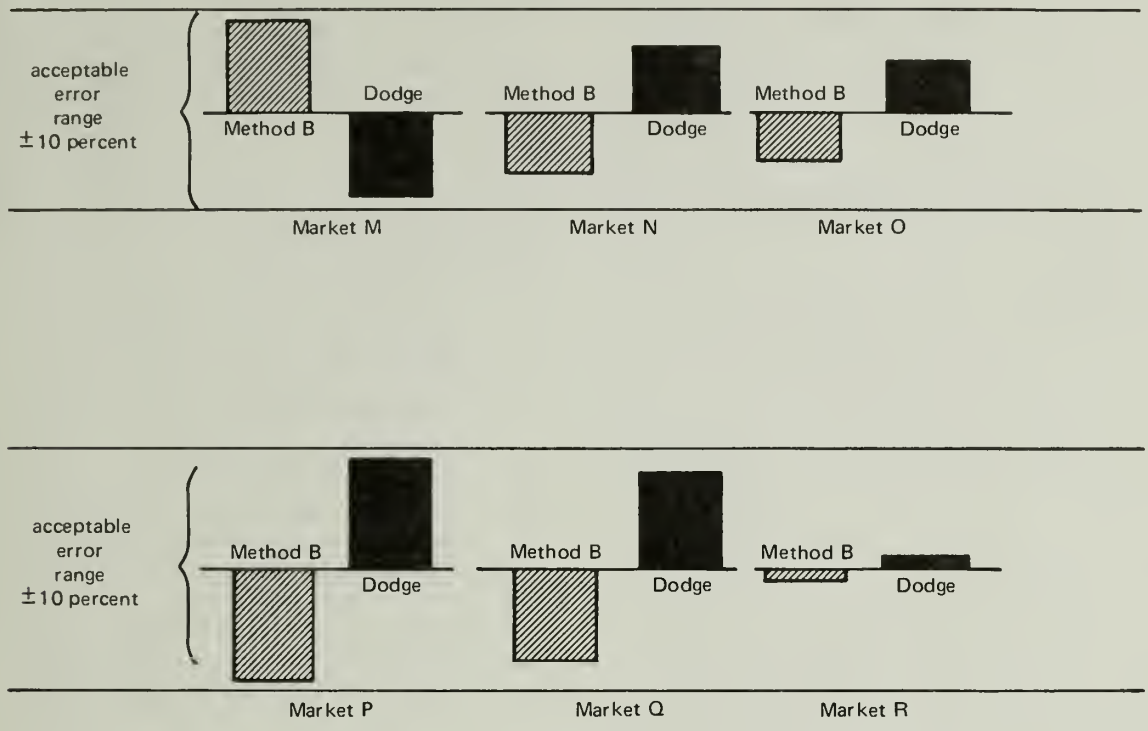


Figure 4. Dodge and Method B Consistency Validation—Medium Markets Six Markets Electrical Contractor Market Potential



comparison can be made for all known trading areas to determine the limits of the null hypothesis.

DISCUSSION

An empirical model was developed for estimating local trading area electrical contractor market potential. The model was validated in 19 different local trading areas. There are a number of approaches for developing empirical models to be used with the Dodge construction statistics data base. Product use factors, extrapolation, correlation, regression, and expert judgment were used in the current application. All of these approaches are based on historical trends and, hence, require periodical updating.

Few sources of information are available for estimating electrical contractor local market potentials. Not even the Bureau of the Census, Census of Construction, provides information which can be accurately applied to local trading areas. The product potential method offers a means to estimate local area market potentials. In addition, the product potential method can be expanded to produce product demand forecasts. (See reference 4.)

REFERENCES

1. King, W.R. Estimating Market Potential. In R. Ferber (Ed.), *Handbook of Market Research*. New York: McGraw-Hill, 1974.

2. King, W.R. Product Testing. In R. Ferber (Ed.), *Handbook of Market Research*. New York: McGraw-Hill, 1974.

3. Morawetz, J.H. Estimating Building Product Demand Through An Input-Output Application. Paper presented at the 39th Session of the International Statistical Institute. Vienna, 1973.

4. Morawetz, J.H. Forecasting Building Materials Demand From General Contracts. Business and Economic Statistical Section of the American Statistical Association. American Statistical Association, 1974.

A Forecasting Model of Regional Housing Construction

Richard S. Conway, Jr.
Washington State Department of Commerce and Economic Development
Charles T. Howard
University of Denver

INTRODUCTION

An alternative subtitle for this paper might well be "Modeling with Weak Theory and No Observations." Apart from the conceptual difficulty of specifying the volatile behavior of residential investment, the task of calibrating a regional forecasting equation is complicated by a scarcity of data. In particular, there are no time-series measurements of either new residential construction-put-in-place or housing stocks for subnational areas. Add to this the common problems of statistical estimation, such as multicollinearity, and one is faced with a number of formidable obstacles. Nevertheless, in spite of such barriers, there is an operational approach to modeling regional housing construction that holds some promise. In the following sections, we demonstrate that a relatively simple theory combined with the efficient use of available data can lead to a reasonable forecasting equation.

THE MODEL

Although there is no universally accepted explanation of residential investment, estimated models at the national level tend to bear a family resemblance, incorporating both demand and supply factors. We postulate a market model derived from three structural equations.[1] The demand equation assumes that per capita demand for housing stock varies directly with per capita income and credit availability and inversely with the price of housing.[2] The supply equation assumes that per capita residential investment (i.e., new construction) is positively related to the housing price and negatively related to construction costs. Finally, the equilibrium condition states that the price of housing adjusts until demand equals supply, when the supply consists of the housing stock at the end of the previous period plus the investment in the current period.

By imposing linear restrictions upon the structural equations, we obtain the following reduced form equation:[3]

$$i_t = B_o + B_1 y_t + B_2 r_t + B_3 c_t + B_4 s_{t-1},\ B_1>0,\ B_2>0,\ B_3<0,\ B_4<0 \quad (1)$$

where:

i_t = per capita investment in housing stock in year t

y_t = per capita income

r_t = a measure of credit availability

c_t = relative construction cost of housing

s_{t-1} = per capita housing stock at end of year t-1.

Equation (1) is not the exact form of the model estimated, since there are no observations on the stock of housing. Following the suggestion of Almon et al. (reference 1), we construct a surrogate variable. Let S_o be the actual *total* housing stock in year zero, I_t be total construction during year t, and 2 percent be the retirement rate. Then the actual stock at the end of year t (S_t) is the sum of the surviving part of structures built since year zero and the surviving part of the initial stock. Hence:

$$S_t = \sum_{j=0}^{t-1} (0.98)^j I_{t-j} + S_o (0.98)^t,\ t\ 1 \quad (2)$$

Putting this on a per capita basis by dividing by the population (n_t) and substituting into equation (1), we obtain:

$$i_t = B_o + B_1 y_t + B_2 r_t + B_3 c_t + B_4 k_{t-1} + B_4 S_o m_{t-1} \quad (3)$$

where

$$k_t = \sum_{j=0}^{t-1} (0.98)^j I_{t-j}/n_t \quad (4)$$

and

$$m_t = (0.98)^t/n_t \quad (5)$$

The validity of the housing model, as expressed by equation (3), can be checked in three ways. First, each coefficient should have the sign stipulated in equation (1). Second, the estimate of the per capita housing stock in the initial period ($\hat{s}_o$), which can be determined from the parameter estimates in equation (3) and the initial period's population, should compare favorably with the corresponding national figure. Third, the estimate of the income elasticity of housing demand ($\hat{E}_y$), which can also be calculated from the coefficients in equation (3) and supporting data, should be in line with previous findings.

THE DATA

It is not an overstatement to say that regional econometrics is largely a data problem. In the case of housing construction,

Note: The authors wish to thank Jonah Otelsberg of the City University of New York for her helpful comments on an earlier version of this paper.

[1] An alternative to the market model is the stock-adjustment formulation of Almon et al. (see reference 1). However, explanatory variables in these two specifications are virtually the same. Only interpretations of the models tend to be different. See also specifications of Muth (reference 6), Preston (reference 7), and Carliner (reference 3).

[2] Inclusion of the credit availability term, instead of an interest rate, is an empirical consideration based on the effectiveness of this variable in previous studies (e.g., Almon et al. (reference 1)).

[3] Signs of the coefficients in equation (1) follow from restrictions on signs of coefficients in the three structural equations.

there exist no State data comparable to the Bureau of the Census Series C-30, Value of New Construction Put in Place, the concept appropriate for our model. However, there are two information sources from which it is possible to make *indirect* measurements: State permit-issuing data from the Bureau of the Census Construction Reports, Series C-40; and State contract construction values reported by F. W. Dodge.[4] After several manipulations, which involve smoothing, scaling, and deflating the data, we obtain two annual series of independent estimates of new residential construction-put-in-place for each State in question. As a step to insure more efficient (in the statistical sense) coefficient estimates, the model is tested with the mean of these two transformed series.

To round out data requirements, we need information on resident population, disposable income (in 1972 dollars), credit, and construction costs. The credit availability term, an indicator of the amount of funds in the mortgage market, is defined as the difference between Moody's AAA corporate bond rate and the interest rate for prime 4-6 month commercial paper. The construction cost variable is the ratio of the Boekh construction cost index to the consumer expenditures deflator.

THE RESULTS

The housing construction equations are estimated using the ordinary least squares (OLS) method on annual observations from 1958 to 1974. The results are found in the table. Shown first is the equation for Washington State, since our research has been conducted with the development of an interindustry econometric model for that region.[5] The housing model has been further tested on data for Arizona, Georgia, Indiana, and New York. Reported in the table are the regression coefficients along with their respective t-values in parentheses. Also given are the corrected coefficient of determination, the Durbin-Watson statistic, and the standard error of the estimate, which in parentheses is expressed as a percentage of the mean value of the dependent variable. Finally, we show the implied estimates of the per capita housing stock for 1957 (in 1972 dollars) and the income elasticity of the demand for housing stock.

In general, the results are encouraging. Note that lagged income and a lagged interest rate differential enter into the final equations, since they outperform their unlagged counterparts. As shown by the regression statistics, the percentage of variation explained by the model is fairly high, there is little evidence of first-order autocorrelation, and the prediction error over the observation period is less than 20 percent in each case.

Of the 25 explanatory variables in the 5 equations, 21 enter with correct signs of the regression coefficients. Regression coefficients with incorrect signs or low t-values are, in part, explained by the high correlation between income and housing cost. For example, when the cost term is dropped from the Georgia model, the estimated equation is more consistent with theory;

$$\underset{}{i_t} = \underset{}{0.6940} + \underset{(2.8)}{0.1741}\,y_{t-1} + \underset{(1.4)}{3.7037}\,r_{t-1} - \underset{(4.5)}{0.1955}\,k_{t-1} - \underset{(-1.6)}{3492.6}\,m_{t-1} \quad \underset{(-2.6)}{}$$

$$\bar{R}^2=0.84,\ DW=1.37,\ SEE=0.0249\ (11.6),\ \hat{s}_o=4740,\ \hat{E}_{y_{t-1}}=0.356$$

As for the estimates of initial stock and income elasticity, both are reasonable in size. The average estimate of per capita housing stock in 1957 for the five States is $6,240, a value slightly above the national figure of $6,100 for the same year. The average income elasticity of 0.407 also compares well with recent findings at the national level. Not all of the previously estimated elasticities are conceptually comparable, which accounts for some of the differences in the estimates. Bearing this in mind, studies based on cross-sectional data by De Leeuw (reference 4), Maisel et al. (reference 5), and Carliner (reference 3) have yielded income elasticities of 1.0, 0.6, and 0.5, respectively, for renters and 1.1, 0.8, and 0.6, respectively, for owners. From their model estimated from times-series information, Almon et al. (reference 1) obtain an income elasticity of 0.3.

As a last demonstration of the model, the housing equation for Washington is graphically depicted in the chart. Despite

[4] We wish to thank the F. W. Dodge Division of the McGraw-Hill Information Systems Company for permission to use proprietary data.

[5] See Bourque, Conway, and Howard (reference 2) for a complete description of the Washington Projection and Simulation Model.

Estimated Regional Housing Investment Equations

Selected States	Intercept	y_{t-1}	r_{t-1}	c_t	k_{t-1}	m_{t-1}	$\bar{R}^2$	DW	SEE	$\hat{s}_o$	$\hat{E}_{y_{t-1}}$
Washington...	0.8919	0.5251	4.7629	-0.8902	-0.3449	-4032.0	0.82	1.87	0.0220	4290	1.004
	(1.3)	(6.2)	(4.3)	(-3.9)	(-4.1)	(-2.3)			(8.9)		
Arizona.....	0.9778	1.1141	10.0640	-2.7262	-0.4511	-1242.5	0.80	1.52	0.0566	2450	1.875
	(1.1)	(1.9)	(4.3)	(-1.1)	(-2.8)	(-1.1)			(16.7)		
Georgia.....	0.8578	-0.2085	0.3223	1.8905	-0.2834	-6768.0	0.89	1.77	0.0207	6340	-0.233
	(3.9)	(-1.1)	(0.2)	(2.5)	(-2.7)	(-3.9)			(9.7)		
Indiana.....	0.3795	-0.0468	1.0493	0.2018	-0.0234	-1244.9	0.91	2.54	0.0061	11750	-0.640
	(1.9)	(-1.7)	(3.8)	(2.4)	(-0.6)	(-1.3)			(3.6)		
New York.....	1.3017	0.0075	2.3011	-0.2383	-0.1549	-16171.1	0.74	0.91	0.0168	6380	0.027
	(1.9)	(0.5)	(3.5)	(-2.5)	(-1.5)	(-1.4)			(10.3)		

wide swings in construction, the model tracks the observation period well, picking up the major turning points. For forecasts outside the sample period, the prediction errors are substantially larger.[6] The ex ante mean absolute percentage error for 1975, 1976, and 1977 is 16.4 percent, which compares with a sample error of 6.5 percent. On the other hand, there is relatively little bias in the 1975-1977 predictions, the difference between the predicted and actual three-year means being only -2.8 percent. Furthermore, a correction for speculative demand considerably improves the fit over the past three years. The recently rapid increase in housing price apparently has triggered a housing demand for investment purposes. As a test of this hypothesis, a speculative demand variable (q_t), defined as the change in housing price relative to the current mortgage rate, has been added to the formulation. Applying the OLS method to annual Washington data from 1958 to 1977 gives the following equation:

$$i_t = 1.1623 + 0.3849y_{t-1} + 2.7934r_{t-1} - 0.5171c_t$$
$$\quad (1.5) \quad (4.4) \quad (3.1) \quad (-1.9)$$
$$+0.5579q_t - 0.3281k_{t-1} - 4399.9m_{t-1}$$
$$\quad (2.7) \quad (-3.1) \quad (-2.1)$$

$$\bar{R}^2 = 0.82,\ DW = 2.32,\ SEE = 0.0298(11.2),\ \hat{s}_o = 4920,\ \hat{E}_{yt-1} = 0.719$$

[6] For the post-observation simulation, per capita income, population, credit availability, and housing cost are assumed to be known. The model therefore predicts not only investment but also the implied housing stock.

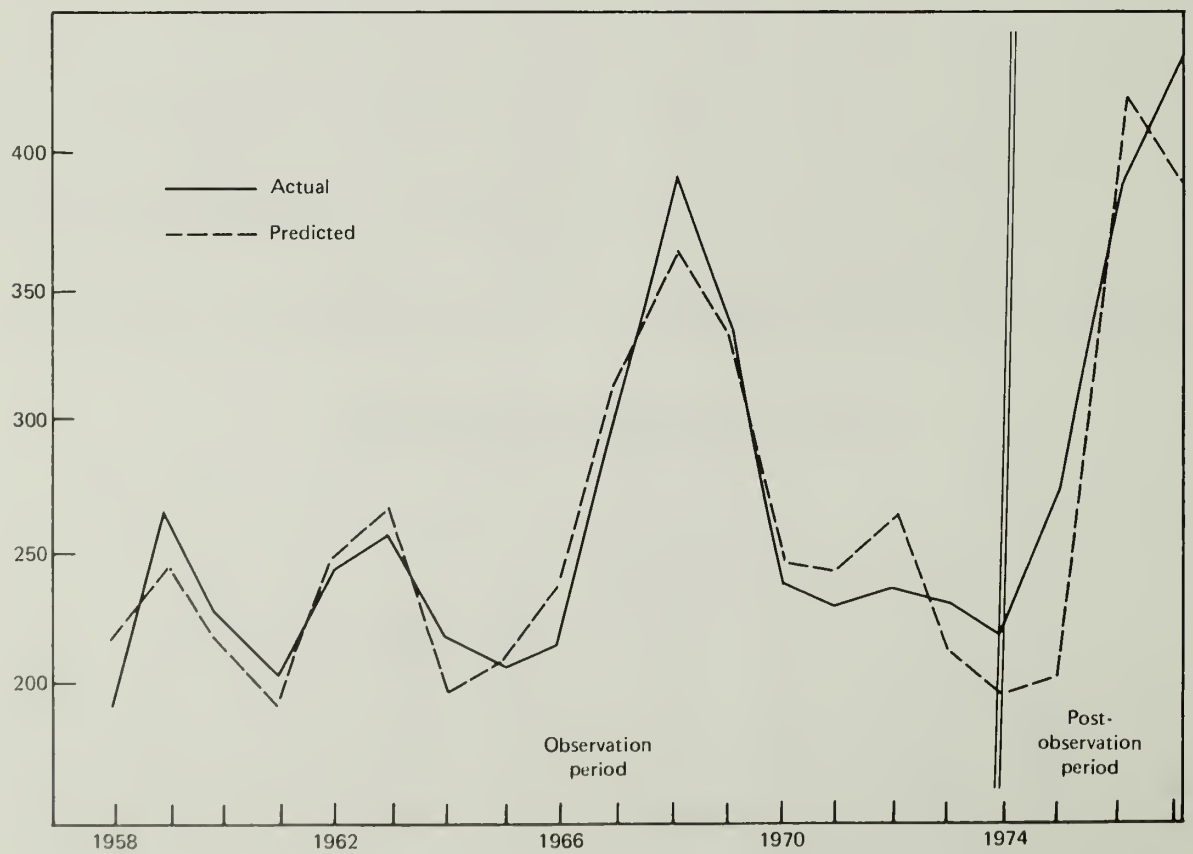
The speculative demand term is not only statistically significant, but its inclusion reduces prediction error over the extended observation period and keeps the model from turning down in 1977.[7]

CONCLUSION

The development of explanatory models of regional growth has been hindered by the dearth of suitable economic information. Indeed, the structures of many small-area models are dictated by data limitations and not by theories of economic behavior, a predicament which often leads to formulations of little analytical value. It is our contention that theory should not be relegated to such a secondary role. What this means is that regional scientists must either build the necessary data base or use available information more efficiently. This study on regional housing investment has adopted the latter approach.

[7] Using another version of the housing model and more complicated variants of q_t, Douglas Pedersen, economist for Rainier National Bank in Washington, has found evidence suggesting that speculative demand has been an even more significant factor in the recent state housing boom.

Per Capita Housing Investment for Washington State
(1972 dollars)



REFERENCES

1. Almon, C., Jr., et al. *1985: Interindustry Forecasts of the American Economy*. Lexington, Mass.: Lexington Books, 1974.

2. Bourque, P. J., R. S. Conway, Jr., and C. T. Howard. *The Washington Projection and Simulation Model.* Seattle: Graduate School of Business Administration, University of Washington, 1977.

3. Carliner, G. "Income Elasticity of Housing Demand," *The Review of Economics and Statistics*, Vol. 55, (1973), pp. 528-532.

4. De Leeuw, F. "The Demand for Housing: A Review of Cross Section Evidence," *The Review of Economics and Statistics*, Vol. 53, (1971), pp. 1-10.

5. Maisel, S. J., J. B. Burnham, and J.S. Austin. "The Demand for Housing: A Comment," *The Review of Economics and Statistics*, Vol. 53, (1971), pp. 410-413.

6. Muth, R. F. "The Demand for Non-Farm Housing," in Harberger (ed.) *The Demand for Durable Goods.* Chicago: University of Chicago Press, 1960.

7. Preston, R. S. *The Wharton Annual and Industry Forecasting Model.* Philadelphia: University of Pennsylvania, 1972.

Discussant

Jonah Otelsberg
City University of New York

Everyone who needs small-area estimates is aware of two major problems:

1. Lack of data for the areas of interest to the researcher.
2. Lack of theory or organized body of knowledge to help with methodology.

The Committee on Small-Area Statistics tries to remedy both problems by encouraging collection of statistics for small areas and providing practitioners with guidance in methodology, using the data available from secondary sources or developing their own sources.

The four papers presented so far demonstrate the thrust of the Committee's activities. The first two papers described data available for small-area analysis. The other two, the ones to which I will confine my remarks, deal with methodology used in obtaining small-area estimates needed for decisionmaking.

Both papers are very well written and presented. Unlike many a discussant, I had copies of the papers well in advance and had the opportunity to study them ahead of time. My thanks to the authors. The most obvious point in both papers is that the data needed for analysis and decisionmaking was not available in exactly the form that people wanted. There is some data available on small areas. However, the definition of small areas is so varied and the needs are so different from one user to another, that it seems there is no substitute, other than providing statistics for very small areas, that would allow researchers to combine these small areas into their areas of interest.

The Committee adopted the definition of small area as an "area smaller than a state." For use by researchers, this should be limited to at least a county. The smaller the unit for which data is available, the easier it is for researchers to obtain data for their needs by aggregation into sales territories or other areas of interest.

Both papers point out that, partly because of the lack of data, a sophisticated method of estimating small-area statistics is not available. Most of the methods used are rather ad hoc, based on the researchers own imagination and ingenuity. Those of you in business or in government who are required to produce data for decisions know that some basis, other than just one's own impression, is better than nothing at all.

As to the two papers, I, myself, am more interested in guidance on methodology than was given in both papers. However, it should be spelled out for researchers step-by-step. In the case of Mr. Ptacek, this information may be proprietory. This is a problem when you deal with private company data, but I hope that some way will be developed to report to colleagues the steps and assumptions being made in producing estimates. I found Mr. Ptacek's paper well designed in terms of validation of the procedures. Very often researchers will produce an estimate and hope for the best. Mr. Ptacek, using three different methods, estimated the criteria variables and validated each one against the others. This is a recommended approach. Of course, it requires resources to do it, but, even if the resources are not available, you can use the recommendation from the first half of this season; that is, obtain estimates for small areas of interest and also independent estimates for aggregates. For example, if county estimates are the purpose of your research, obtain county estimates, but also obtain independent State estimates and compare the sum of the counties to the States as the minimum validation of the performance of your model.

Mr. Conway, in his paper, did not have outside validation of the model. The graph at the end of his presentation showed the actual versus predicted values as a result of the regression equation based on data that was used in developing the equation itself. This shows the adequacy of the data for the period on which the regression equation was based, but it does not justify using the regression estimates for any other periods. I was happy to hear at the presentation that the following periods were looked at. A good way to solve this problem is to leave out some parts of available data for the purpose of validation, rather than wait for the time when the equation can be tested because new data became available.

Also, in the Conway paper, all the regression coefficients for the State of Washington were significant. Equations for other States have many regression coefficients which are not significant. This is a problem in selecting a particular equation. Economists tend to stay with the models they selected and get a predictive equation: an equation that includes all the variables they have selected regardless of significance of coefficients. A more efficient, pragmatic approach may be to exclude all those variables that have an insignificant regression coefficient and to fit new equations to just those variables that are better predictors without dependence upon the theory. The theory is weaker that way, but the estimates might be better.

After recently giving a seminar in Research Methodology, I have found that the simple truths are very often forgotten. Let me repeat them again. Define your problem carefully: the criteria measurement, the hypothesis, the assumptions have to be spelled out. Don't try to reinvent the wheel. Search the literature for work by others in your field. Speak to colleagues. If it is a situation that involved proprietory data, articles may not be published on the work done. Speak to others in your field and see who has done work in your area. It is most likely that they will be willing to share their ideas and experiences with you. Ascertain which data is available, both secondary sources and administrative records. Usually, within any organization, there is a great deal of data available in administrative records, although administrative records are not a very good source for research. However, if you have no other source, and, if you have studied the limitations of that administrative record, it may be very useful. If you have the choice; that is, if you have the funds, develop your own sources. Even a small survey (not on the scale that Mr. Ptacek had) to validate your data may be very useful. State your hypothesis precisely. Define criteria for

testing your hypotheses and the validity of your results, and spell out all the assumptions. There are implicit assumptions made when selecting a method. For example, when you use a regression equation, there is an implicit assumption of normality; there is an implicit assumption that the errors are distributed with zero mean and a known variance. Think about whether that model fits your data. Spell it out. It's better to err by saying too much instead of saying too little. Validate your estimates by using a different method (as Mr. Ptacek did) or develop your own ways. As I said earlier, a good way of doing it might be getting an estimate for areas for which data are available; getting national, regional, or State estimates, which would be sums of the local estimate that you may keep, and evaluate your local estimates in that way.

Small-Area Estimators: County Crop Acreage Estimates Using LANDSAT Data

Manual Cárdenas
New Mexico State University
Michael E. Craig and Mark Blanchard
Department of Agriculture

This research considers several county estimators which incorporate LANDSAT satellite data with that obtained from USDA's operational June Enumerative Survey (JES). The radiometric satellite data are classified into the different crop types using a maximum likelihood discriminant function. The classified data is then used as an auxiliary variable to JES questionnaire data. Approximate variance formulas for the proposed county estimators are presented.

INTRODUCTION

This paper deals with the estimation of small-area characteristics from a sample designed for making large-area estimates. Particular interest is given to making crop acreage estimates at the county level from data obtained in the June Enumerative Survey (JES), a survey conducted at the State and national levels.

The Economics, Statistics, and Cooperatives Service (ESCS) has been charged with making area estimates of crops based on the JES. County estimates are an integral part of the ESCS program of crop estimates. ESCS receives direct funding for making certain county estimates and has annual agreements with Agriculture Stabilization and Conservation Service (ASCS) and the Federal Crop Insurance Corporation to provide selected additional county data. State statistical offices (SSO's) are responsible for the preparation of county estimates. The county estimates are made by first allocating the official State estimate for a given crop proportionately among crop reporting districts (collections of contiguous counties) and then, apportioning the estimates for these districts among the individual counties. Besides the information obtained from the JES, the SSO's also use data in their estimation procedures derived from several other sources. Two such sources are: A mail survey which may include 50 to 100 respondents; and the agricultural census. The estimation procedure thus varies from State to State and from county to county depending upon the availability of data. No variance estimates are computed, but the coefficients of variation are believed to be on the order of 10 percent or more.

Note: Manuel Cárdenas is a 1977-78 faculty member with the Statistical Research Division, Room 4844 South Building, Washington, D.C. 20050.

Since the advent of LANDSAT data, the New Techniques Section of the Statistical Research Division (SRD) of ESCS has focused its resources on the development of methodology that incorporates these data with that obtained from the JES for more efficient estimation. The potential for efficient estimation, as well as a uniform county estimation procedure using LANDSAT data, has been recognized and is presently being investigated.

Actually, the small-area estimation problem has attracted considerable attention in other governmental agencies as well. The National Center for Health Statistics (references 5 and 6) and the Department of Commerce (reference 3), for example, are involved in developing small-area estimators for certain characteristics (e.g., unemployment rates, percent of population having completed college, percentage disabled by chronic conditions, population growth, etc.) from large-area samples such as the Current Population Survey (CPS) and the Health Interview Survey.

DATA ACQUISITION

Before proceeding to the estimators, a brief discussion to acquaint the reader on the data acquisition seems imperative. A more detailed discussion can be found in several sources (e.g., see references 4 and 9).

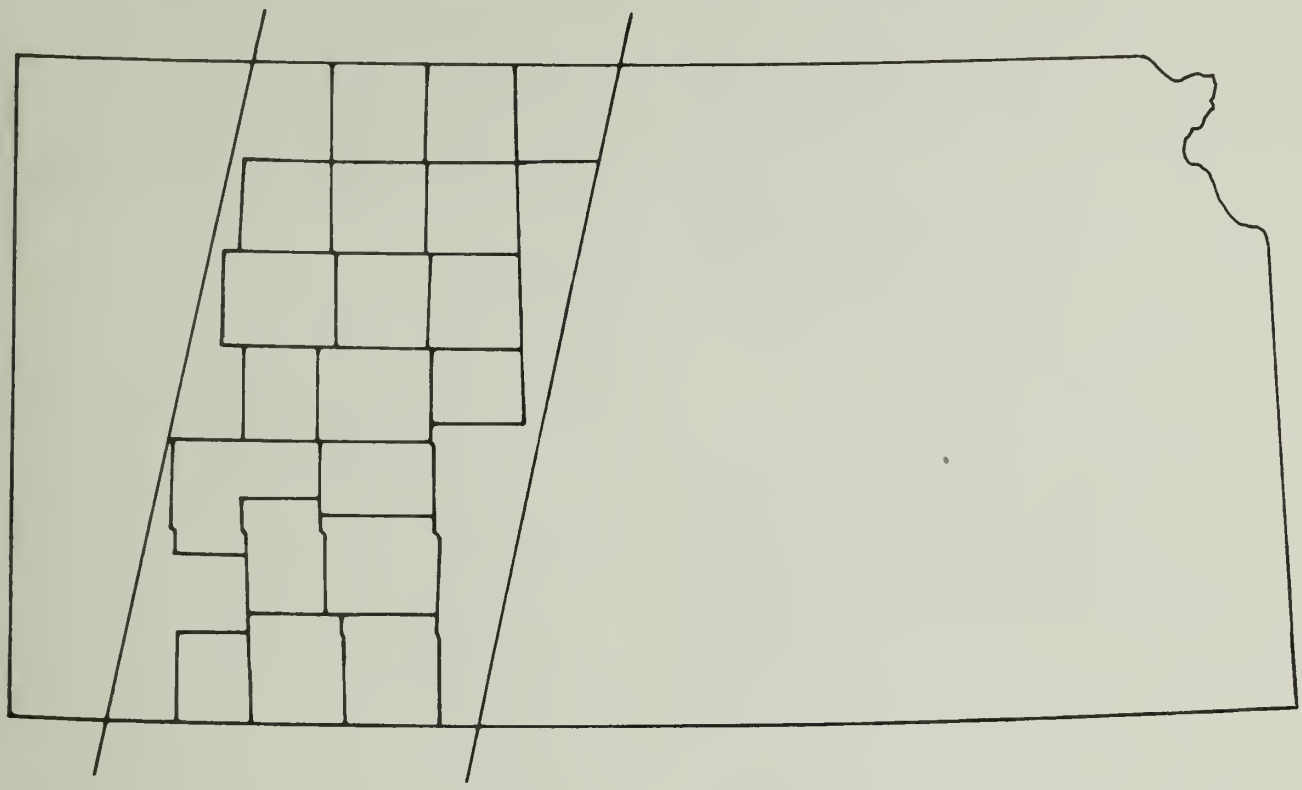
The JES is an annual agricultural survey conducted in late May. The sample for this survey employs two levels of stratification. The first level strata are the 50 individual States. The secondary strata are areas within a State which have similar land use patterns as determined by photo-interpretation of aerial photography. The secondary strata are divided into primary sampling units which can be further subdivided into sampling units. The sampling units chosen for the JES are called segments and are well-defined areas of land varying in size depending on the stratum in which they are located. Typically, these segments are one-square mile in size in the more cultivated strata. The acreage devoted to each crop or land-use are recorded for each field in each segment during the JES interviews.

The basic element of LANDSAT data is called a signature and is the set of measurements taken by the satellite's multispectral scanner (MSS) of an area of the earth's surface approximately one acre in size. The individual MSS resolution areas are called pixels. The MSS measures the amount of radiant energy reflected and/or emitted from the earth's surface in various regions (bands) of the electromagnetic spectrum.

Presently, satellite data is obtained from LANDSAT II and LANDSAT III. A given point on the earth's surface is imaged once every 18 days by the same satellite and once every 9 days by either of the 2 satellites. Each satellite pass covers an area 185 kilometers wide. Figure 1 shows one such pass over the State of Kansas.

The satellite information used by ESCS is extracted from LANDSAT data by classifying individual signatures as to probable crop type. This classification is performed by a collection of discriminant functions. Therefore, LANDSAT data is census data but of questionable reliability due to misclassification.

Figure 1. LANDSAT Pass Wholly Containing 19 Kansas Counties



PRELIMINARY DISCUSSION

The county estimation procedure presented here assumes that the mean number of pixels per segment in stratum h within county i classified as the crop in question, $\overline{X}_{ih}$, is fixed with respect to the JES sample. With the present procedure of sampling and classification, this assumption is not satisfied. However, with a large enough sample, the variability of these values should be negligible in comparison with the variability of the y_{ijh} values (i.e., the reported acreage of the crop in question in the j^{th} segment of the h^{th} stratum within the i^{th} county). A recent study (reference 7), using a jackknife method on 83 sampled segments, tends to verify this.

In developing the estimates, the JES data which was taken at the segment level must be combined with the LANDSAT data which can be taken at the county level. This is done by noting that, whenever a segment is chosen, the county in which that segment is contained is automatically selected also. Moreover, taking a small sample without replacement from a large population is practically equivalent to taking the sample with replacement from that population. To the extent that these two procedures of sampling are the same, it can be seen that taking a simple random sample of n segments from a State is the same as the following two-stage sampling scheme:

(a) A sample of n counties is taken with replacement and with probability proportional to size.

(b) A simple random sample of t_i (t_i being the number of times county i appears in the sample) segments is taken from each of the distinct counties in the sample.

This two-stage sampling procedure was first proposed in a more general form (i.e., a subsample of size $m_i t_i$ rather than t_i is taken from the i^{th} primary unit in the sample) by Sukhatme and Sukhatme (reference 8). The estimators and variances presented in this paper are based on this two-stage sampling scheme. The derivations of variances and their estimators follow the logic used by Sukhatme and Sukhatme and are found in reference 1.

COUNTY ESTIMATORS

If the assumption were made that the mean per segment in land-use stratum h of the crop in question for each county were equal to the mean of the populations $\bar{\bar{Y}}_h$, the total for a particular county, say county k, would be

$$Y_k = \sum_{h \in C_k} M_{kh} \bar{\bar{Y}}_h$$

where $\sum_{h \in C_k}$ denotes the summation over all strata in county k,

and M_{kh} = total number of segments in the h^{th} stratum within the k^{th} county.

An unbiased estimate of Y_k is

$$\hat{Y}_k = \sum_{h \in C_k} M_{kh} \bar{\bar{Y}}^*_h$$

where $\bar{\bar{Y}}^*_h = \frac{1}{n_h} \sum_{i=1}^{N_h} t_{ih} \bar{Y}^*_{ih}$ - an unbiased estimate of $\bar{\bar{Y}}_h$;

$\bar{Y}^*_{ih} = \sum_{j=1}^{t_{ih}} y_{ijh} / t_{ih}$ - the sample mean of the acreage per segment in stratum h within county i;

n_h = number of counties (distinct or otherwise) in the sample of the h^{th} stratum,

and N_h = number of counties containing any part of the h^{th} stratum.

Recognizing that the above assumption is not satisfactory in general, we then search for supplementary information which indicates deviation of a particular county mean from the population mean. This information is found in the form of classified pixels in each county. Using these auxiliary data, we define the family of estimators,

$$\hat{Y}_{Bk} = \sum_{h \in C_k} M_{kh} [\bar{\bar{Y}}^*_h + B_h (\bar{X}_{kh} - \bar{X}_h)] \qquad (1)$$

where $\bar{X}_h$ = the mean number of pixels classified as the crop in question for stratum h. If $\bar{X}_{kh}$ is greater (less) than the mean of stratum h for the given satellite pass, then the mean area estimate should be increased (decreased) by an amount proportional to this difference. It follows that the B_h's should be positive.

If classification is such that $y_{ijh} = A x_{ijh}$, where A is some constant, then using $B_h = \bar{\bar{Y}}^*_h / \bar{X}_h$ in equation (1) yields an unbiased estimator, $\hat{Y}_{rk}$, of Y_k. Other possible values which one might try for the B_h's would be the least squares-like estimates.

$$B_h = \frac{M_h \sum_{i=1}^{N_h} t_{ih} (\bar{X}_{ih} - \bar{X}_h) \bar{Y}^*_{ih}}{n_h \sum_{i=1}^{N_h} M_{ih} (\bar{X}_{ih} - \bar{X}_h)^2}$$

These values of B_h substituted in (1) yield unbiased estimates, $\hat{Y}_{sk}$, of Y_k when $y_{ijh} = a + b_h x_{ijh}$, where a and b_h are constants. Actually, in this case B_h is an unbiased estimate of $C ov (\bar{X}_{ih}, \bar{Y}_{ih}) / V (\bar{X}_{ih})$ for all h. If $b_h = b$ for all h, then we can use the combined data for all strata to estimate b. In this case, substitution of

$$B_h = \frac{\sum_{h=1}^{L} \frac{M_h^2}{n_h} \sum_{i=1}^{N_h} t_{ih} (\bar{X}_{ih} - \bar{X}_h) \bar{Y}^*_{ih}}{\sum_{h=1}^{L} M_h \sum_{i=1}^{N_h} M_{ih} (\bar{X}_{ih} - \bar{X}_h)^2},$$

where L is the number of strata, gives unbiased estimates, $\hat{Y}_{ck}$, of Y_k. The sum over k for all three of the estimators, $\hat{Y}_{rk}$, $\hat{Y}_{sk}$ and $\hat{Y}_{ck}$, is unbiased for the population total.

The estimators, $\hat{Y}_{rk}$ and $\hat{Y}_{sk}$ can be written as

$$\tilde{Y}_k = \sum_{h \in C_k} M_{kh} \left[\frac{1}{n_h} \sum_{i=1}^{N_h} w_{ih(k)} \, t_{ih} \, \bar{Y}^*_{ih}\right]$$

where

$$w_{ih(k)} = \begin{cases} \bar{X}_{kh} / \bar{X}_h, & \text{for } \hat{Y}_{rk} \\ 1 + M_h \dfrac{(\bar{X}_{ih} - \bar{X}_h)(\bar{X}_{kh} - \bar{X}_h)}{\sum_{i=1}^{N_h} M_{ih} (\bar{X}_{ih} - \bar{X}_h)^2}, & \text{for } \hat{Y}_{sk} \end{cases}$$

The estimator, $\hat{Y}_{ck}$, can be written as

$$\hat{Y}_{ck} = \sum_{\ell \in C_k} M_{k\ell} \sum_{h=1}^{L} \left[\frac{1}{n_h} \sum_{i=1}^{N_h} w_{ih\ell(k)} t_{ih} \bar{Y}^*_{ih}\right]$$

with

$$w_{ih\ell(k)} = \delta_{\ell h} + \frac{M_h^2 (\bar{X}_{ih} - \bar{X}_h)(\bar{X}_{k\ell} - \bar{X}_\ell)}{\sum_{h=1}^{L} M_h \sum_{i=1}^{N_h} M_{ih} (\bar{X}_{ih} - \bar{X}_h)^2}$$

and $\delta_{\ell h} = \begin{cases} 1 \text{ if } \ell = h \\ 0 \text{ otherwise} \end{cases}$

This estimator will not be discussed further since its variance should be, at best, as large as the variance of $\hat{Y}_{sk}$.

The variance for $\tilde{Y}_k$ is derived in formula (1) and is given by

$$\begin{aligned} V(\tilde{Y}_k) = \sum_{h \in C_k} M_{kh}^2 \Bigg\{ & \frac{1}{n_h} \sum_{i=1}^{N_h} (M_{ih}/M_h)[w_{ih(k)} \bar{Y}^*_{ih} \\ & - \sum_{i=1}^{N_h} \frac{M_{ih}}{M_h} w_{ih(k)} \bar{Y}^*_{ih}]^2 \\ & + \frac{1}{n_h M_h} \sum_{i=1}^{N_h} (M_{ih} - 1) w_{ih(k)}^2 S_{ih}^2 \\ & - \frac{n_h - 1}{n_h M_h^2} \sum_{i=1}^{N_h} M_{ih} w_{ih(k)}^2 S_{ih}^2 \Bigg\} \end{aligned} \qquad (3)$$

where $S_{ih}^2 = \dfrac{\sum_{j=1}^{M_{ih}} (y_{ijh} - \bar{Y}_{ih})^2}{M_{ih} - 1}$

and $\bar{Y}_{ih} = \sum_{j=1}^{M_{ih}} y_{ijh}/M_{ih}$. The variance for $\hat{Y}_{rk}$ and $\hat{Y}_{sk}$ are obtained from formula (3) by the appropriate substitution for $w_{ih(k)}$.

If the assumption is made that the within-county variance is equal for all counties, then an unbiased estimate of the variance formula given by (3) is

$$v(\tilde{\bar{Y}}_k) = \sum_{h \epsilon C_k} M_{kh}^2 [n_h(n_h - 1)]^{-1} \left\{ \sum_{i=1}^{n_h} (w_{ih(k)} \bar{Y}^*_{ih} - \frac{1}{n_h} \sum_{i=1}^{n_h} w_{ih(k)} \bar{Y}^*_{ih})^2 + s^2_{wh} \left[\sum_{i=1}^{n_h} (1 - 1/t_{ih}) w^2_{ih(k)} - \frac{n_h - 1}{M_h} \sum_{i=1}^{n_h} w^2_{ih(k)} \right] \right\} \quad (4)$$

where

$$s^2_{wh} = \frac{\sum_{i=1}^{n'_h} \sum_{j=1}^{t_{ih}} (y_{ijh} - \bar{Y}^*_{ih})^2}{n_h - n'_h}$$

is the pooled within-county variance and n'_h = the number of distinct counties in the sample within the h^{th} stratum.

Again, estimated variances for $\hat{Y}_{rk}$ and $\hat{Y}_{sk}$ are obtained by the appropriate substitution for $w_{ih(k)}$. The assumption of equal-within-county variances is needed because some counties have only one observation in some strata. Actually, in most cases, it takes more than one pass of the satellite to completely cover a State. Since these passes occur at different dates and since signatures for the same crop differ from pass to pass, each pass is used as a post stratum. The county estimation is therefore made by post strata which relaxes the assumption from equal-within-county variances for the State to equal-within-county variance within each pass.

CONCLUSIONS

This estimation procedure was tried by the New Techniques Section of ESCS on 40 percent of the Kansas 1976 JES winter wheat data (reference 2). The results seem promising, but, unfortunately, they can only be compared to the SSO estimates which are of unknown reliability. Presently, the procedures are being tried on the 1978 JES data for Iowa.

As was mentioned in the text, the estimators suggested are unbiased under certain linear conditions. However, the classification is not strictly linear. The classification and, therefore, the estimation is expected to get better when LANDSAT D data become available in 1981.

One could also consider other values for the B_h's. Also, a regular regression estimator could be developed. This approach would require "super-population" considerations.

REFERENCES

1. Cárdenas, M., M. Blanchard, and M. E. Craig, "On the Development of Small Area Estimators Using LANDSAT Data as Auxiliary Information," ESCS, U.S. Department of Agriculture, Washington, D.C., 1978.

2. Craig, M. E., R. S. Sigman, and M. Cardenas, "Area Estimates by LANDSAT: Kansas 1976 Winter Wheat," ESCS, U.S. Department of Agriculture, Washington, D.C., 1978.

3. Gonzalez, Maria Elena, and Joseph Wakeberg, "Estimation of the Error of Synthetic Estimates," unpublished paper presented at the First Meeting of the International Association of Survey Statisticians, Vienna, Austria, 1973.

4. "June 1978 Enumerative & Multiple Frame Survey," Interviewer's Manual, ESCS, U.S. Department of Agriculture, Washington, D.C.

5. National Center for Health Statistics, "State Estimates of Disability and Utilization of Medical Services: United States, 1969-71," DHEW Publication No. (HRA) 77-1241, Washington, D.C., 1977.

6. Schaible, W. L., R. J. Cassidy, G. A. Schnack, and D. B. Brock, "Small Area Estimation: An Empirical Comparison of Conventional and Synthetic Estimators for States," submitted for publication.

7. Sigman, R. S., C. P. Gleason, G. A. Hanuschak, and R. R. Starbuck, "Stratified Acreage Estimates in the Illinois Crop-Acreage Experiment," Proceedings of the 1977 Symposium on Machine Processing of Remotely Sensed Data," Purdue University, W. Lafayette, Indiana, 1977.

8. Sukhatme, P. V., and B. V. Sukhatme, *Sampling Theory of Surveys with Applications.* Iowa State University Press, Ames, Iowa. 1970.

9. Von Steen, D. H., and W. H. Wigton, "Crop Identification and Acreage Measurement Utilizing LANDSAT Imagery," Statistical Reporting Service, U.S. Department of Agriculture, Washington, D.C., 1976.

1977 Economic Censuses and Their Use in Private and Public Sectors

Introduction

Edward J. Spar
Marketing Statistics

Every five years, the Bureau of the Census takes an economic census. Relatively little publicity is given to this operation compared to the decennial census. The latest economic census (1977) has been more overshadowed due to its proximity to 1980. If you look at the American Statistical Association agenda this year, you will note that this is the only session dealing directly with the economic census. Considering the importance of these collections to the business community and local governments, this is surprising.

It is impossible to plan production, develop marketing strategies, set sales goals, allocate capital resources to new facilities without knowing about the flow of goods and services throughout the country. Local governments cannot possibly measure the health of their areas or make decisions on how to allocate their resources without having data on the business activity of their community.

The economic census covers many aspects of the economy including transportation, services, manufacturing, wholesale and retail trade. This session will confine itself to one of these areas, retail trade. The purposes of this session are threefold. First, what are the activities of the Bureau of the Census in this area? What are they doing to collect meaningful data and how are they putting it in the most usable form. Second, how are these data actually being put to use? And third, what is being done in the private sector to update this information? From this, we hope to have a clear picture of the importance of this data base.

Meeting Users' Needs from the 1977 Economic Censuses From the Data Collector's Point of View

Shirley Kallek
Bureau of the Census

INTRODUCTION

This quotation from a paper concerning the statistical needs of American business given at the annual meeting of the American Statistical Association 60 years ago is just as applicable today as it was in 1918.

"That, in the future, statistical requirements will increase rather than diminish seems certain. The past decade has witnessed many changes in the political, social, and industrial life of the nation; but the next few years promise even greater, perhaps revolutionary changes. Statistics will be required as never before by those in high places, both in business and in governmental affairs, as a guide for the right solution of the questions of the day. The statistician will be called upon to furnish these data properly analyzed and correlated."[1]

This early evaluation of the future trend of statistical data needs has not only been confirmed by the experience of the last six decades but it also appears to be an accurate forecast of developments in the foreseeable future.

It would seem that users' needs for data are seldom completely satisfied, whereas, the data collector's ability to satisfy these needs usually fall short of the mark. Nevertheless, continuing dialogue between the data user and the data collector is necessary to achieve better and more meaningful figures for both the public and private sectors. The economic censuses whose content is so fundamental to the Government's economic statistics program are probably one of the best examples of the beneficial results of such an interchange.

HISTORY OF THE CENSUSES

The economic censuses provide a rich body of statistical data, which, when examined over time, reflect the economic concerns of each period. It is clear, then, that useful and productive interchanges between users and collector's must have been occurring since the earliest censuses. The history of the censuses is a long one, going back, in the case of manufactures, to 1810. Data for the mineral industries were first collected in 1840. The distributive trades and services censuses were conducted at irregular intervals between 1929 and 1954, although some data were collected as part of the decennial censuses in the early 1900's.

Illustrative of efforts to provide data, in response to the needs of the times, are the major data expansions introduced in the 1880 Census of Manufactures. To describe the changes brought about by the Industrial Revolution, some 49 specialized questionnaires were designed asking specific questions about different manufacturing industries. A Standard Industrial Classification (SIC) system did not exist. In addition, a special survey on wages and prices was included as well as special inquiries on labor activity and trade associations.[2] The expansion in the coverage of the economic censuses in the 19th century paralleled the growth of the American economy. Questions asked also reflect the customs of the time. For example, the 1910 Census of Manufactures collected data on the number of wage earners under 16, while the size classes for number of hours worked per week included a category for 72 hours and over per week. Collection efforts in the 1900's through the World War II period were quite modest, particularly with regard to the general operating characteristics of establishments. With the passage of the Full Employment Act of 1948, the emerging influence of fiscal policy and the more precise measurement of the various phases of the business cycle utilizing the estimates of the Gross National Product (GNP), the demands made by government and private policymakers for more detailed information increased significantly and has continued up to the present time.

At the time statistical needs were expanding, significant advances were being made in statistical methodology. Research found that the utilization of administrative records of other government agencies could materially facilitate the conduct of a census program by providing more accurate mailing lists as well as basic data about individual firms. The computer, of course, was a major new resource. These breakthroughs greatly facilitated the development of an integrated census program where the various economic censuses, previously conducted on an individual basis using both mail and personal interview techniques, could be collected completely by mail from one list and on a uniform basis. The 1954 Economic Censuses were the first to be conducted on this integrated basis. With the inclusion of transportation in 1963 and the reinstitution of the census of construction in 1967, the economic censuses now include all establishments classified in manufacturing, mining, retail trade, wholesale trade, services, construction, and selected areas in transportation. The major expansion in industry coverage for services for the 1977 program will be discussed.

Why continue to conduct a complete census on such a wide scale, at periodic intervals in today's environment, when sampling techniques have been developed to such a fine degree? We believe that only by a complete count of all industries covered in the census at periodic intervals can the necessary detail and flexibility be obtained.

The main advantage of a census comes from the fact that there is a separate report for each establishment classified in the industries covered by the census. The establishment becomes the building block; by assigning a 4-digit SIC code and a physical location code to each establishment, the resulting

[1] Hathaway, William A., "Internal and External Statistical Needs of American Business," American Statistical Association Quarterly Journal, New Series No. 122, June 1918.

[2] U.S. Bureau of the Census, Economic Censuses of the United States: Historical Development, Charles G. Langham, Working Paper No. 38, Washington, D.C., U.S. Government Printing Office, 1973.

data can be examined and tabulated in a myriad of ways. For example, one can study industry changes by size and types of establishment; look at legal form of organization; determine the kinds of products made or sold by an industry; look at the inputs of labor, materials, and capital consumed by an industry and relate these inputs to the outputs of that industry. A census also provides the data base for rebenchmarking many of the interim current economic statistical series. We believe that a complete census now is more important than ever before. However, our definition of a complete census is changing. We collect reports from large firms and use administrative records for smaller ones. Some questions are collected only on a sample basis. But the important thing is that the basic data are available for every establishment within the census scope.

HOW DATA PRIORITIES ARE SET

The data content of any census must reflect the current and future needs of data users in both the public and the private sectors. The Authority for Census, Title 13, United States Code is purposefully vague on the content of the census program. Section 131 of title 13, directs the Secretary of Commerce to undertake censuses of manufactures, of mineral industries, and of other businesses, including the distributive trades, service establishments, and transportation for the years ending in 2 and 7. This responsibility is delegated to the Director of the Census Bureau. Section 5 gives the Secretary of Commerce authority to determine the content of the statistical inquiries. This permits the flexibility necessary in designing a program which is responsive to the needs at a given time without having to revise the census law for each round of censuses.

The Census Bureau acts as a filtering agent in determining the content of a program. In practice, the final content of any census program results from a distillation of the ideas, suggestions, and expressed needs of the many relevant groups. For example, from the data collector's point of view, several things are taken as certainty. First, there will always be far more requests for new and expanded data items than can possibly be met through the censuses. Second, is the continuing changes in our complex society will be accompanied by requests for more detailed information at greater levels of accuracy and timeliness. The data collector faces a challenge of attempting to meet these data needs in the face of ever stronger constraints.

These restrictions on the data collector are extremely significant because they conflict directly with satisfying data requests. The factors must be weighed against the need for information and must be evaluated in some consistent manner so that the requests can be reduced to some manageable number. Major constraints include budgetary limitations, total reporting burden, reportability of the specific information, and the need to achieve an acceptable level of reliability in the resulting data. Because of the magnitude of the economic censuses, which now include over 6 million establishments, these constraints become even more significant.

Minimizing respondent burden must be given serious consideration. The data collector must obtain the cooperation of respondents regardless of the mandatory status of the survey if accurate and meaningful data are to be obtained. A standard axiom should be that reliable data cannot be obtained through coercion. Thus, the inquiry on the horsepower of motors was dropped from the 1963 Census of Manufactures because it was found to create an undue reporting burden and reasonably accurate data could only be obtained at substantial cost to the respondent. Data on manufacturers' sales by class of customer are collected only once every 10 years, although, data is needed more frequently by the Bureau of Economic Analysis. Most manufacturers cannot report the information without undue burden.

How does the data collector first find out what is needed and then determine the priorities for meeting these needs? In discussing the approach used to evaluate data requests, I should like to separate the data content into two parts because the Bureau treats each quite differently. The first group relates to the detailed products and the specialized inquiries for particular industries. The second relates to either the broader issues such as questions on general operating characteristics which are asked of all establishments in the particular census or to the expansion of industry coverage for the census.

Regarding the first group, far more changes are made to the product and specialized industry inquiries in each census than to the inquiries relating to general operating characteristics. Requests for changes in product information come primarily from discussions with trade associations, business firms, and similar sources. These requests have an impact, generally, only on one industry and come primarily from data users who are also the respondents. These requests are accumulated during the intercensal period, although, much of the product detail is collected in the Current Industry Reports series (CIR), many of the changes are introduced annually. Changes in product definitions and content are made to reflect market changes, new technology, and the rise and fall of specific industries.

Experience, over the years, has also shown that product information is much more readily available from company records and that it is much easier for the Bureau to ascertain if the information is readily available from most company records. Finally, it is much easier to develop standardized dollar criteria for breaking out or combining product lines.

Since 1975, an added constraint in changes to product data exists since the classification systems of imports, exports, and domestic production have been significantly revised to achieve greater comparability among the three systems. Any changes proposed in the detail for domestic production data must be reviewed for consistency against the import and export commodity program.

It is much more difficult to determine the basis for deciding priorities among general operating statistics inquiries. These data concern general statistics such as employment, payroll, inventories, etc., and are collected from all establishments in each census. Information collected in this area has been fairly consistent over the past 25 years and the type of detail is shown in table 1. The Bureau of the Census has not been able to quantify in any reliable manner all the variables involved and develop a mathematical model which would rank the re-

Table 1. 1977 Economic Census Check List of Data Items

Item	Mining	Manufactures	Wholesale trade	Retail trade	Selected services industries	Construction industries
Employment:						
Production (construction workers, quarterly)	X	X				X
Other employees	X	X				X
Total employment, quarterly			X	X	X	X
Total employment, annual average	X	X				
Payrolls:						
Production workers (construction)	X	X				X
First quarter total	X	X	X	X	X	X
Annual total	X	X	X	X	X	X
Supplemental labor costs:						
Legally required programs	X	X	S	S	S	X
Other programs	X	X	S	S	S	X
Total	X	X	S	S	S	X
Production workers work-hours, quarterly	X	X				
Total receipts	X	X	X	X	[1]X	X
Inventories:						
By stage of fabrication		X				
End of 1976 total	X	X	X	S		
End of 1977 total	X	X	X	S		
Method of valuation	X	X	X	S		
Operating expenses:						
Cost of electricity, purchased	X	X	S	S	S	X
Products bought and resold	X	X				
Cost of fuels consumed	X	X	S	S	S	X
Advertising			S	S	S	
Rental payments:						
Buildings and structures	X	X	S	S	S	X
Machinery and equipment	X	X	S	S	S	X
Total	X	X	S	S	S	
Cost of materials	X	X		S	S	X
Office supplies			S			
Containers and packaging			S			
Communications	X	X	S	S	S	X
Purchased repairs:						
Buildings and structures	X	X	S	S	S	X
Machinery and equipment	X	X	S	S	S	X
Total	X	X	S	S	S	
Contract work	X	X				X
Total operating expenses			X			
Fixed assets:						
Beginning of year:						
Buildings and structures	X	X	S	S	S	
Machinery and equipment	X	X	S	S	S	
Total	X	X	S	S	S	
Acquisitions			S	S	S	
Deductions	X	X	S	S	S	
Depreciation:						
Buildings and structures	X	X	S	S	S	X
Machinery and equipment	X	X	S	S	S	X
Total	X	X	S	S	S	X

See footnote at end of table.

Table 1. **1977 Economic Census Check List of Data Items**—Continued

Item	Mining	Manufactures	Wholesale trade	Retail trade	Selected services industries	Construction industries
Fixed assets--Continued						
Capital expenditures:						
New buildings and structures	X	X	S	S	S	X
New machinery and equipment:						
Transportation equipment			S	S	S	X
Computers and related equipment			S	S	S	
Other machinery, equipment			S	S	S	X
Total	X	X	S	S	S	X
Used:						
Buildings and structures						X
Machinery and equipment						X
Total	X	X	S	S	S	X
End of year:						
Buildings and structures	X	X	S	S	S	X
Machinery and equipment	X	X	S	S	S	X
Total	X	X	S	S	S	X
Quantity of electricity:						
Purchased	X	X				X
Generated	X	X				
Sold	X	X				
Legal form of organization	X	X	X	X	X	X

[1]Total expenses collected for nonprofit service industries.

quests in priority order to produce cost-benefit analyses or a scoring system as a means of selection among the numerous requests for new information. To a large extent, the final decision of what will be collected is arrived at after long and lengthy discussions with industry and government agencies.

It is important to have a coordinated and systematic plan for obtaining the views of data users and of respondents who must supply the information. While it may seem easy in theory, it becomes difficult in practice. We, of course, attempt to utilize all avenues. The Bureau analysts in each of the subject matter divisions have frequent contacts with data users and there is a constant interchange of ideas relating to new data needs. The problem is to assimilate these requests and assemble them in a systematic manner prior to any census review. In an initial step in preparing for the 1977 Economic Censuses, we called together an ad hoc committee of knowledgeable individuals from universities, private business, and government to review the general framework of the economic censuses program and to suggest areas where expansion was needed, where concepts were outdated, or which needed more detailed study in light of today's economy. For example, the whole area of inventory data with its inconsistencies and its problems was discussed.

In addition, we took tremendous advantage of the formal mechanisms set up through the Office of Federal Statistical Policy and Standards. An interagency committee of about 20 government agencies was established and meetings were held with the group as a whole as well as on individual agency basis. Each user agency also submitted a detailed list of their requirements. Meetings were held concurrently with congressional committees and with private groups such as the Business Advisory Council on Federal reports, trade associations, and business firms. The Bureau's advisory committees were also consulted. The data user conferences held after the 1972 Economic Censuses were also important sources of information. Much attention was also given to the recommendations of the GNP Improvement Committee which had been established by the Office of Management and Budget.

Thus, the data requests come from a multitude of sources but all go through the same filtering process. As part of the 1977 program, a major decision was made to conduct a recordkeeping practices survey to determine the reportability of the new item requests. The last recordkeeping practices survey had been conducted as part of the 1958 censuses and had covered only about 100 large manufacturing firms. As a Bureau paper on the recordkeeping practices survey pointed out—since that time, (1958), while company structure had become far more complex, the implementation of electronic data processing equipment had made more information more accessible. In this way, one could at least determine the response burden and reportability of the new information requests and establish a rational basis for weeding out selected inquiries.[3] The survey

[3]Kallek, S. and L.H. Lyons, "Determining the Data Content for Economic Censuses," 1977 Business and Economic Statistics Section Proceedings of the American Statistical Association, pp. 234-238.

was also designed to give the Bureau an opportunity to review the reportability and burden of the general statistics items which had been collected in previous censuses. This survey was conducted in the latter part of 1976 and provided an important input in eliminating those items which could not be collected in a satisfactory manner or which would result in undue burden. The remaining items were then reviewed to determine the geographic level at which the data were required. Were the data needed at a local area level, or would national and regional totals suffice? If the latter were true, the new data items were considered for inclusion only on a sample basis. I might add that sampling was used to a much greater extent in the 1977 program than ever before, because, in addition to reducing the cost of collecting the information, there is also a significant reduction in reporting burden. This is particularly important since most of the new items under consideration were more difficult to collect.

After the general statistics data requests were reviewed for reportability and costed out, an attempt was made to assign priorities to the requests. Some of the factors included an assessment of the need for the data: (1) by Government economic policymakers; (2) for improvement of the GNP accounts; (3) for productivity measures; and (4) for business planning and marketing activities. The process seems to work because the continuing dialogue brings about a consensus of opinion. All inquiries of course were cleared through the Office of Management and Budget assisted by the Business Advisory Council on Federal reports. Table 2 summarizes the new items included in the 1977 program as a result of this process. The major expansion came in the service areas where the nonprofit sector was included for the first time. The other major changes were primarily in the collection of additional data items relating to inventories and assets.

OTHER AREAS OF CHANGE

Another challenge faced by the data collector is to make the data more useful by retabulation or by making minor additions to the report form. Significantly, more information can be obtained in this way and this has been done in several areas for 1977.

An example was the inclusion of a check-box inquiry in 1977 to permit respondents to self-designate themselves as department stores. Up to now, the major problem in developing information for discount stores was the lack of concensus on a uniform definition. We hope that the use of a check-box inquiry where companies can designate themselves as a "discount department store" will resolve this problem. The data and characteristics of those designating themselves in this manner will be analyzed and compared with those that do not. We believe that we can obtain reasonable results in this manner and publish figures on employment, payrolls, floor space, etc., for this segment of department stores.

The Commodity Transportation Survey has been revised to link the shipments reported in this survey directly to the total shipments reported in the census of manufactures. This will permit a better understanding of the relative uses of the definite modes of transportation and the origins and destinations of the manufacturing sector.

While, the major retail center program for retail has been limited in scope, a major advance is a joint effort with cities of over 500,000 in population. This city economic area program will result in tabulations of census data for the various economic censuses by subcity areas. These subcity areas have been determined by planning groups within the cities and the program will be done on a reimbursable basis. To date, 25 cities have decided to enter this program.

DISSEMINATION OF RESULTS

From the Bureau's point of view, the collection of reliable, meaningful information is just one phase of its responsibility. Disseminating the results effectively is equally important. We are deeply interested not only in providing users with earlier access to the census results, but in providing the data in alternative forms to the publication of hard copy.

The 1977 census results will begin to be issued in published reports in December 1978 with the start of the release of preliminary or advance State and industry reports. The greater bulk of the final reports will be issued in the latter part of 1979 and early 1980. The detailed publication schedule for hard copy and microfilm will be released within the next several months. We plan to use microfiche and microfilm concurrently with the printed reports to a much greater extent than ever before, as well as issuing summary tapes as was done for the 1972 census. This time, however, the tapes and microfiche will contain additional data not available in the printed reports. The series of data user conferences will be expanded and a new data user conference program has been developed for selected user groups such as small business owners, etc. Much of the census results will also be highlighted in an economic census atlas which will be published for the first time. Needless to say, the Census Bureau always stands ready to provide special tabulations on a reimbursable basis.

CONCLUSION

The way the Census Bureau views the economic censuses from a data collector's point of view are: The many varied needs of all data users are assessed in a systematic manner to assure that the data collected are statistically valid and reportable without undue burden; the total program fits within the budgetary limitations set by the Executive Branch and the Congress; and the statistics are disseminated in ways that result in maximum utilization by data users.

The 1977 Economic Censuses are extremely broad in scope. Almost 400 different questionnaires are needed to survey the range of economic activity in the United States. This complexity makes it difficult to remember exactly what will be asked in any specific census. The following table has been drawn up to help determine, in advance, what can be expected to flow

from this important canvass of the Nation's economic structure. An "X" indicates that questions on identified line item will be asked in a given census, and an "S" indicates that for retail trade, merchant wholesalers, and selected services, data will be collected on a sample basis and results will be published on a national level only.

Table 2. **Economic Censuses New Data Inquiries**

Mineral Industries (all establishments)

1. Purchased communications services.
2. Rental payments for buildings and structures.
3. Rental payments for machinery and equipment.
4. Inventories.
5. Breakout of capital expenditures for used buildings and structures separate from used machinery and equipment.
6. Depreciation charges for the year.
7. Retirements from fixed asset accounts during the year.

Construction Industries (all establishments)

1. Employer costs of fringe benefits.
2. Purchased communications services.
3. Purchased repairs to buildings and structures.
4. Purchased repairs to machinery and equipment.
5. Rental payments for buildings and structures.
6. Purchased fuels and electric energy.
7. New capital expenditures for automobiles and other transportation equipment.

Manufactures

All Establishments

1. Salaries and wages of employees engaged in transportation for the company's account—to be collected for a few SIC's only.
2. Salaries and wages of employees engaged in construction for the company's account—to be collected for a few SIC's only.
3. New capital expenditures for equipment broken out between "purchased for own use" and "purchased for rental to others"—on form for selected SIC's only.

On Sample Basis Only

1. Purchased communications services.
2. Purchased repairs to buildings and structures.
3. Purchased repairs to machinery and equipment.
4. Breakout of capital expenditures for used buildings and structures separate from used machinery and equipment.
5. New capital expenditures for transportation equipment.
6. New capital expenditures for computers and related equipment.
7. Depreciation charges for the year.
8. Retirements from fixed asset accounts during the year.

Retail Trade

All Establishments

1. Second, third, and fourth quarter employment.
2. Method of inventory valuation (tire, battery, and accessory stores only).
3. Selected questions to identify specific kind of business:
   a. Home centers
   b. Discount department stores
   c. Truck stops
   d. Fast food operations
   e. Antique stores
   f. Pawn shops
   g. Convenience food stores
   h. Catalog showrooms
   i. Furniture warehouse showrooms
   j. Membership organizations
   k. Ophthamologists
4. Parts installed in repair work and service labor charges (selected kind of business).
5. Additional broad merchandise line.

On Sample Basis Only

1. Salaries and wages of employees engaged in transportation for the company's account.
2. Purchased communication services.
3. Purchased repairs to buildings and structures.
4. Purchased repairs to machinery and equipment.
5. Purchased fuels and electric energy.
6. Purchased advertising services.
7. Cost of purchased materials and supplies.
8. New capital expenditures for transportation equipment.
9. New capital expenditures for computers and related equipment.
10. Depreciation charges for the year.

Wholesale Trade

All Types of Operation—On Establishment Basis

1. Uniform commodity line inquiries for all types of operation.
2. Method of inventory valuation for all classifications.
3. Intracompany transfers.

Table 2. Economic Censuses New Data Inquiries—Continued

Wholesale Trade—Continued

All Types of Operation—On Establishment Basis—Continued

4. Second, third, and fourth quarter employment.
5. Employment by principal activity (merchants).
6. Credit sales, receivables, and bad debt losses (sales branches and offices).
7. Service receipts and labor charges.

Merchant Wholesalers on Sample Basis Only

1. Salaries and wages of employees engaged in transportation for the company's account.
2. Purchased communications services.
3. Purchased repairs to buildings and structures.
4. Purchased repairs to machinery and equipment.
5. Purchased fuels and electric energy.
6. Purchased advertising services.
7. New capital expenditures for transportation equipment.
8. New capital expenditures for computers and related equipment.
9. Method used for inventory valuation—to be collected on establishment basis for all wholesalers.

Service Industries

All Establishments

1. Expanded scope (SIC 702, 704, 80 ex. 8072, 82, 83, 84, 86, ex. 8661, 89 ex. 8911):

a. Receipts for taxable operations and expenses for tax-exempt operations.
b. Annual and first quarter payroll and employment for first through fourth quarter.
c. Depreciation charges, fixed assets, and capital expenditures of tax-exempt organizations.
d. Reimbursable payroll expenses of expense-sharing organizations.
e. Selected sources of revenue of tax-exempt organizations.

Service Industries—Continued

All Establishments—Continued

f. Ownership or control by a religious organization of tax-exempt organization.
g. Annual payroll and number of personnel by occupational class for physicians, dentists, hospitals, nursing care facilities, educational institutions, and accounting firms.
h. Primary field of practice for physicians.
i. Selected facility characteristics of hospitals.
j. Program specialty of social services.
k. Analysis of revenue of membership organizations and tax-exempt social service organizations.
l. Analysis of receipts of accounting firms.

2. "Old" service (SIC 70-79 ex. 702, 704; 8072, 81 and 8911):

a. Second, third, and fourth quarter employment.
b. Sales of merchandise (all SIC's).
c. Capital expenditures for new machinery and equipment (computer and data processing services, automobile and truck rental, leasing establishments, and equipment rental and leasing.

On Sample Basis Only

1. Purchased communication services.
2. Purchased repairs to buildings and structures.
3. Purchased repairs to machinery and equipment.
4. Purchased fuels and electric energy.
5. Purchased advertising services.
6. Cost of purchased materials and supplies.
7. New capital expenditures for equipment broken out between "purchased for own use" and "purchased for rental to others"—on establishment forms for selected SIC's only.
8. Depreciation charges for the year.
9. New capital expenditures for transportation equipment.
10. New capital expenditures for computers and related equipment.

Methodologies Used for Updating Retail Trade Data From the Economic Censuses

Elias Fokas
Market Statistics

Although the 5-year economic census fulfils the needs of many sections of our national economy, there are a lot of industries which depend on the most current information to evaluate their marketing systems and to plan a certain line of action for the immediate future. Pure population trends are not enough anymore to determine how much people will spend in retail stores or in what kind of stores they spend their money. The fast food establishments have shifted tremendously the money spent in eating and drinking places in the last few years, and prescription medicines can now be purchased in many more department stores as opposed to a few years ago. Examples like the above can be numerous since the nature of retail stores is changing. Hence, the need of updated information on the retail activity of a county or other geographical area.

Both of the approaches I'm going to discuss have the following similarities:

a) Assume the benchmark to be the last economic census of retail trade for the 10 major store groups on county level;

b) the creation of updated state control totals; and

c) utilize published information which can be purchased at relatively low cost from Federal, State, or local governmental agencies, as opposed to expensive and time-consuming, individual, market-sampling approaches.

THE CREATION OF UPDATED STATE TOTALS

Since the final updated estimates at the State and national level must agree with officially accepted estimates, the **Monthly Retail Trade** reports can be used to create State totals for total retail sales and the 10 major store categories. I will take it for granted that everyone in this room is familiar with the above publication from the Department of Commerce and, if that's the case, you will know that the December issue has the annual summaries. In this issue, besides the national estimates, you will find complete distributions for the 4 geographical regions and the 9 census divisions, as well as the 15 most industrialized States. Therefore, you can form nine matrices, one for each geographic division. Fill in the States for which the **Monthly Retail Trade** report published estimates. For the undisclosed States, apply to the benchmark the annual growth of retail taxes collected by the State as published in the state tax collections reports. If you can't find any information for a particular State, then you can use annualized growth between the last two censuses. At this point in our matrix, there should be a preliminary estimate for every State within the geographic division. All that remains to be done is to control the undisclosed States to the balance of the division. Here, the balance of the division is defined as the divisional totals minus the disclosed States, and, by controlling, we mean that the summation of the undisclosed States in the division must be made equal to the balance of the division. This controlling can be done by proportionally adjusting the first State estimates.

Having the State control totals, you can now go to the county level. As the heading indicates, the first approach is using retail taxes collected at the county level. There are about 45 States which collect and publish information on retail sales taxes, and these reports can very easily be obtained by the State Department of Revenue. These reports are inexpensive and come on a monthly, quarterly, or semiannual basis. The way to use the tax collections is by measuring the percent change from one year to the next and applying the growth to last year's retail sales for the store groups. The sum of the stores then will provide an estimate for total retail sales for the county. The final step is to control the summations of the counties to the previously arrived State totals. By the mere fact of what they represent, the taxes collected seem to me the best indication of the retail trade activity of the county. The store, by law, has to charge sales tax on every sale, making the taxes directly proportional to the actual sales of the store. It could, at this point, be asked: Why aren't the sales calculated by dividing the amount of taxes by the given tax rate? That would be the ideal situation if the classification of the store by the State were exactly the same as by the Bureau of the Census. However, this is not the case. By expressing the taxes as a growth over last year's, you eliminate the misclassification of the stores. Besides the possible tax rate inconsistency throughout the year, another shortcoming of the approach is the underreporting of taxes by the stores. The procedure goes around this problem by assuming that the underreporting is consistent from county to county and the final proportional controlling to the State total serves that purpose as well. It becomes obvious now how important the State control totals are, and how careful one has to be in making certain that the collected local taxes are adjusted for any tax rate changes through the year.

The second approach in updating retail sales at the county level is based on county business patterns. This approach attempts to measure the retail trade activity of the county based on the assumption that employment shifts from one kind of establishment to another or from a store of a certain employment size to another will reposition the county in relation to the State for that store group. The basic idea here is to redefine the employment in a certain store group to reflect the relative importance of an employee, in terms of sales, according to the size of store he is working in. This relative importance of the employee can be calculated using a special tabulation from the 1972 Census of Retail Trade, Series RC72-S-1, table 1b, which gives you retail sales by establishment size as well as the number of employees by establishment size. Therefore, you can calculate sales per employee by kind of store and by size of establishment. If you now consider sales per employee in the low employment size (which is 1 to 4 employees) having an

index of 1, you can calculate the indices of each higher group by dividing the sales per employee for each of these groups by the sales per employee for the lowest group. By multiplying each of the individual indices by the average number of employees in each employment size, you will come up with a set of weights, by kind of store, which represents at the national level the relative importance, in terms of sales, of each employee according to the size of store he works in. With these weights, you can to to the County Business Patterns publications for year 1972 and create for each store in each county a quantity which can be called effective retail trade employment and can be defined as the sum of the cross products of number of establishments in size i by the national weight for size i (as described above). Express that quantity as percent of State and associate that percent with the percent of actual sales of county to State in the 1972 Census of Retail Trade. To update from this point on, calculate the 1973 effective retail trade employment as percent of State and apply the 1973/1972 change to the 1972 census percent of actual sales of county to State. The result of this multiplication is the projected 1973 percent of actual sales of county to State and, since you have already created State control totals by store, the final step would be to multiply projected 1973 percent actual sales of county by the State total to arrive at the projected 1973 sales of the county for that store group. Repeating the process for each one of the 10 major stores, you will have a complete county distribution of retail sales by type of store, and the summation of the major stores is the county's total retail sales.

To go over one simplified example, consider the following:

The Creation of U.S. Weights by Size of Establishment for Food Stores

Controls	Employment size					
	1 to 4	5 to 9	10 to 19	20 to 49	50 to 99	100 or more
Sales per employee.....(dollars)..	54,360	40,714	48,614	61,712	61,219	56,093
Relative index.....	1.0000	.7490	8943	1.1352	1.1262	1.0319
Average number of employees.....	2.1	6.9	14.6	32.1	66.6	150.0
Weight.....	2.1	5.2	13.1	36.4	75.0	154.7

Note: Special tabulations taken from 1972 Census of Retail Trade.

$$\text{Effective food employment} = \sum_{j=1}^{6} \text{1972 County Business Patterns (number of establishments in size j)} \times \text{(1972 census weight for store of size j)}$$

for county A

j = 1 (Size 1-4)
j = 2 (Size 5-9)
j = 6 (Size 100+)

$$\text{then } \left(\frac{\text{1973 effective food employment county A\% of State}}{\text{1972 effective food employment county A\% of State}}\right) \times \text{(1972 census food sales county A\% State)} = \text{1973 county A food sales \% of State}$$

and

1973 county A food sales = "3" x 1973 State food sales

Again, you can see the importance of good, updated, State control totals.

The Application of Economic Censuses Data in Analyzing the Food-Away-From-Home Market

Malcolm M. Knapp
Malcolm M. Knapp, Inc.

The purpose of my paper is to discuss how to analyze the food-away-from-home market, also known as the food service industry. In the course of the discussion, I will focus on what constitutes the food service industry; the application of economic census data to it; and the reliability of other available data (using grades of soft, medium, and hard data). This will bring us to the central themes of the paper, which are:

1. The economic censuses are the *central control points* for all research in the food service industry except for research in schools, medicine, some government programs, some employee feeding, and other institutional populations. Of the total sales volume of the food service industry, 83.15 percent is based in some major way on the economic censuses.
2. Be suspicious of all data on the food service industry. There are so many contradictions about the size of different market segments and so many revisions that you must make careful comparisons before choosing which data to use.
3. Organize and modify the data to fit the requirements of your user group.
4. Use your judgement. Don't be reluctant to challenge even official data if they don't make sense.

My working definition of the food service industry is that it encompasses all meals, snacks, and drinks *prepared outside* the home. Thus, takeout meals and beverages are included in the food service industry. A picnic or a brown bag lunch prepared inside the home for consumption outside the home is not part of the food service industry. My key distinction is where the food is prepared, not where it is eaten.

An annual report that I prepare for the National Restaurant Association divides the food service industry into three major groupings which contain a total of 54 separate market segments. This level of detail is made possible in large part because of the merchandise line data of the economic census. The three major groupings are:

1. *Commercial.* This major group comprises those establishments which are open to the public, operated for profit, and may operate facilities and/or supply meal service on a regular basis for others. It is the largest of the segments, accounting for 84.46 percent of the total industry sales, excluding the military segments.

2. *Institutional feeding.* The second group comprises business, educational, government, or institutional organizations which operate their own foodservice. Food is provided as an auxiliary service to complement their other activities. While some establishments operate at a profit, this is not the aim of the food service activity. Rather, they serve food principally as a convenience for their own employees, students, patients, etc. Note that I include the contract feeders as part of the commercial group, even though they operate many of their facilities in what I describe as the institutional feeding group. This is because they are in business to make a profit. Since each market segment is broken out within the contractor group, the data can be easily combined to provide totals by pure market segment. While the institutional feeding group has only 15.54 percent of the sales, it has 26.16 percent of the total purchases, excluding military. This is because I count only actual sales or imputed sales where the consumer has paid for a total service.

3. *Military feeding.* I treat this group separately because most suppliers of food, etc. treat it as a distinct entity for sales purposes.

A good question at this point is: Why have 54 separate segments? Why not have 10 or 12 as some other food service industry analysts use. This is a very complex industry which contains several communities of industry data users. My approach, therefore, is to break the industry into as many fine segments as the data support so that any industry participant can recombine the fine segments into different major groups or subgroups to suit the particular purpose at hand. User categories rarely have sharp, neat boundaries.

Another set of reasons is that the usage rates of a specific product per dollar of purchases can vary substantially by fine market segment. Fast food (limited menu restaurant) have a lower usage rate of coffee than cafeterias. As an example, to group all SIC 5812 Eating Places together in one lump is just not valid. Not only are there differential usage rates, there are differential basic growth rates as well as differential distributions of establishment size and chain-ownership concentration. These factors have a significant bearing on hurdle volumes for key account sales calls. So that you understand what all these segments are, I will read the list quickly.

As is apparent to any user of the various economic censuses, much of the list is taken from detail provided in these censuses. The most important data are retail trade, merchandise line sales, miscellaneous subjects, selected services, and hotels, motels, trailering parks and camps. The first question is: What do we want to include in our sales measurement? The best vehicle is the merchandise line sales data. Because we—more importantly, the purveyors to the industry—want to know the sales volume of food and drink. The disposable container manufacturers are interested in the volume of takeout business. Through the mechanism of the merchandise line sales data, this specific information can be provided. I use the merchandise line sales data for all the base data on sales and establishment count for the segments covered by retail trade. For the main channels of distribution, such as restaurants, I calculate the sales volume not covered by the merchandise line sales data due to noncoverage

The Food Service Industry—Estimated Food and Drink Sales and Purchases: 1976

Type of establishment	Number of units	Estimated food and drug sales (thousand dollars)	Total food and drug sales (percent)	Estimated food and drug purchases (thousand dollars)	Total food and drug purchases (percent)	Source
Grand total	(X)	78,908,418	(X)	34,644,505	(X)	(X)
Group I Commercial Feeding[1]						
Total, group I	(X)	66,163,875	84.46	23,279,370	73.84	(X)
Restaurants, lunchrooms	[2]112,180	26,541,072	33.88	10,539,791	33.43	1,21,22,23
Social caterers	3,944	1,037,451	1.33	421,464	1.34	1,22
Commercial cafeterias	8,222	2,499,595	3.19	948,122	3.01	1,22,23
Limited menu restaurants (refreshment places)	80,609	14,903,154	19.03	5,360,366	17.00	1,21,22,23
Ice cream, frozen custard stands	5,550	551,025	0.70	187,348	0.59	1,21,22
Bars and taverns	[3]44,112	6,308,491	8.05	[4]385,953	1.22	1,22
Food Contractors: Total	[5]5,836	(X)	(X)	(X)	(X)	(X)
Manufacturing, industrial plants		1,450,683	1.85	676,018	2.14	1,19,27
Commercial and office bldgs	(X)	379,368	0.48	176,786	0.56	1,27
Hospitals and nursing homes	(X)	565,293	0.72	226,117	0.72	13
Colleges and universities	(X)	755,783	0.97	267,540	0.85	4
Primary and secondary schools	(X)	448,177	0.57	210,643	0.67	3
In-transit feeding (airlines)	(X)	301,041	0.38	[6]144,500	0.46	9
Recreation and sports center	(X)	603,967	0.77	223,468	0.71	1,23,24,29
Hotel restaurants	13,438	2,463,415	3.15	826,175	2.62	1,23,25,26
Motor hotel restaurants	2,498	608,650	0.78	204,806	0.65	1,23,25,26
Motel restaurants	13,551	1,367,656	1.75	496,994	1.57	1,26
Drug and prop. store restaurants	9,323	489,114	0.62	185,863	0.59	1
Gen. merchandise store restaurants	1,269	38,996	0.05	14,818	0.05	1
Department store restaurants	3,882	809,163	1.03	323,665	1.03	1,30
Variety store restaurants	6,509	542,180	0.69	211,449	0.67	1
Food stores, except grocery	3,299	157,717	0.20	53,624	0.17	1
Grocery store restaurants	12,579	360,272	0.46	133,301	0.42	1
Gasoline service stations	7,738	173,731	0.22	64,280	0.20	1
Drive-in movies	3,384	107,097	0.14	35,342	0.11	1
Misc. retailers (liquor, cigar, etc.)[7]	3,622	127,372	0.16	46,491	0.15	1
Vending and nonstore retailers[8]	2,750	1,500,306	1.92	510,104	1.62	1,19
Mobile caterers	(X)	295,017	0.38	103,256	0.33	1,2
Bowling lanes	3,866	329,840	0.42	135,234	0.43	1
Recreation and sports centers	(X)	448,249	0.57	165,852	0.53	1,24,29

[1]Data are given only for establishments with payroll. [2]Figures are latest Bureau of the Census area reports or merchandise line detail counts or updates when reliable data become available. [3]Unit count includes only those establishments serving food; however, sales figure is for all bars and taverns with payroll. [4]Food only. Cost of alcoholic beverages totaled $1,914,312,000. [5]Individual businesses, not locations. Contract feeders are included in eating place totals in all Bureau of the Census publications although their sales volume figures for contract feeders are significantly understated. [6]Food purchases only. [7]Includes SIC 59, except 591 and 596. [8]Includes sales of hot food, sandwiches, pastries, coffee, and other hot beverages.

The Food Service Industry—Estimated Food and Drink Sales and Purchases: 1976—Continued

Type of establishment	Number of units	Estimated food and drug sales (thousand dollars)	Total food and drug sales (percent)	Estimated food and drug purchases (thousand dollars)	Total food and drug purchases (percent)	Source
Group II Institutional Feeding--Business, Educational, Government or Institutional Organizations Which Operate Their Own Food Service						
Total, group II.....	(X)	12,170,540	15.54	8,249,570	26.16	(X)
Employee Feeding:						
Industrial and commercial organizations.....	4,000	955,050	1.22	467,065	1.48	20,27
Sea-going ships (1,000+ tons)...	548	43,866	0.06	26,320	0.08	6
Inland waterway vessels.....	4,248	133,909	0.17	81,348	0.26	7
Public and parochial elementary and secondary schools (89,381)...	92,297					
National school lunch program[1]....	(X)	1,734,482	2.22	2,337,845	7.41	3
Colleges and Universities:[2]						
Public.....	980	1,133,325	1.45	646,947	2.05	4
Private.....	1,407	450,069	0.57	257,196	0.82	4
Transportation:						
Passenger/cargo liners.....	61	71,650	0.09	39,407	0.12	8
Airlines.....	32	284,744	0.36	141,838	0.45	9
Railroads.....	2	22,512	0.03	14,828	0.05	10
Clubs.....	10,310	737,929	0.94	355,298	1.13	5
Voluntary proprietary hospitals...	4,120	3,223,485	4.12	1,289,394	4.09	13
State, local short-term hospitals[3]	1,836	481,640	0.61	346,789	1.10	13
Long-term general, TB, nervous and mental hospitals.....	746	733,707	0.94	293,482	0.93	13
Federal hospitals[3].....	380	215,354	0.27	190,229	0.60	13
Nursing homes, homes for aged, blind, orphans, mentally and physically handicapped[4].....	26,672	1,638,299	2.09	1,052,214	3.34	14
Sporting and recreational camps...	3,165	85,769	0.11	51,461	0.16	1,11
Community centers.....	16,010	224,750	0.29	265,198	0.84	12
Convents and seminaries.....	(X)	([5])	(X)	109,123	0.35	17
Penal Institutions:						
Federal and state prisons.....	620	([5])	(X)	156,976	0.50	15
Jails.....	3,921	([5])	(X)	126,612	0.40	16
Food furnished food service employees in groups I and II.....	(X)	(X)	(X)	2,099,741	(X)	18
Group III Military Feeding						
Total, group III.....	(X)	574,003	(X)	1,015,824	(X)	(X)
Defense personnel.....	(X)	(X)	(X)	799,527	(X)	18
Officers and NCO clubs ("open mess")[6].....	(X)	370,992	(X)	126,972	(X)	28
Food service--Military exchanges[6].	(X)	203,011	(X)	89,325	(X)	28

X Not applicable.

[1]School lunch program commodities furnished in the calendar year 1976 under Sec. 6,32,416, are worth $473,224,487. In addition, 2,282,051,651 half pints of milk worth $147,786,139 were supplied to 83,555 outlets. [2]Total number of colleges and universities which have food service whether contracted or not. [3]Represents only sales or commercial equivalent to employees. [4]Sales (commercial equivalent) calculated for nursing homes and homes for aged only. All others in this grouping make no charge for food served either in cash or in kind. [5]These institutions make no charge for food. [6]Continental United States only.

of nonpayroll establishments. This is an allocation, since fine segment data is revealed only for establishments with payroll. The merchandise line data mix of meals and snacks vs. alcoholic beverages is critical to calculating the dollar volume of purchases, as the alcoholic beverage sales percent of total receipts varies considerably from segment to segment. I use the economic censuses to establish growth rates by segment to aid in making annual estimates.

Using data from the economic censuses is not without its problems. There are classification problems in this very dynamic industry which cause severe difficulties in making comparisons from one census to another. For example, establishing restaurant guidelines is not an easy matter and is subject to change over a period of time. The census officials have a policy, which I endorse, of making the most accurate and reasonable definition possible at the time of the taking of the census. Thus, in the 1977 census, in addition to the 1972 criterion, an eating place establishment must have waiter or waitress service while the patron is seated in order to be a restaurant. This was adopted to get around the problem of limited menu establishments which had high check averages but which were rapidly being classified as fast-food places. Also, as the industry evolved, a criterion was needed which all could recognize. This placed an establishment such as Pizza Hut into the restaurant category in 1977. In 1972 it was probably classified as a fast-food refreshment place. Indeed, during that time, Pizza Hut was evolving into a family restaurant. What the statistics will show is a loss of the Pizza Hut volume from fast food and a gain by restaurants. While this is, in fact, true because of the evolution, the establishments in question did not go out of business or come into business, as the raw figures might suggest. I will return to the classification problem a little later with an attempt at a solution.

There are several segments for which the retail trade census estimates seem at variance with other data. Census estimates are too low by a big margin in the food contractor area. There is also some concern about the size of the vending and nonstore retailer segment.

There are two other areas for concern. One is that a big push is being made this time (1977 census) not to record sales tax values. Census officials know that some establishments did include sales tax in their 1972 figures. In preparing historical revisions, the current retail trade statisticians took several percent off the 1972 Census of Retail Trade figures for eating places to adjust for the sales tax.

The other area is the way the establishments are being counted. The net effect of the procedure change is that small restaurants which have common ownership will be grouped as one establishment. This will be an undercounting of the establishments, particularly if one makes comparisons to the 1972 census. However, census officials say they will present figures which will blow the establishment counts up to comparable terms with the 1972 data. A person unfamiliar with the data could draw erroneous conclusions.

Returning to the classification issue in the eating places group, the data users complained that the data weren't broken down fine enough for their sales purposes on the part of manufacturers or for competitive analysis on the part of restauranteurs.

My response to the classification problem was to look for some common criteria which were few in number, easily understood, and did a reasonably good job of creating mutually exclusive and collectively exhaustive categories. I have already stated that the boundaries between segments aren't sharp and neat. So I call my effort at classification "Spectrum." This analogy to the world of physics seems to provide a good physical analog, because the color is pure in the middle of its band and then there is a fuzzy area before you get to the next color. Restaurant classification is similar in that the paradigm restaurant in a category is, by definition, pure, and, as you go toward the next category, things aren't so pure and, in fact, can and do get fuzzy.

The classification areas are menu (very limited, limited, full, luxury), price (low, low-moderate, moderate, moderate-high, high), type of service (takeout, snack stand, self-service, service, continental service) and, as a final qualifier, level and type of decor.

The research was sponsored by Restaurant Business magazine and is called Restaurant Business Spectrum. The final form of the classification effort is this fold-out chart. It contains 11 main groupings of bands of restaurants. Each band is characterized by which selection of point on the three criteria has been made. For example, one group is limited menu, low price, self-service and is further qualified by the name of the paradigm, McDonald's. It turned out to be very important to include the paradigm name because users of information in the industry were knowledgeable about certain restaurants which have received a good deal of publicity. This helps establish a mental image of the restaurant type in question.

For the 11 categories, I show sales, number of units, food sales, takeout sales, takeout units, alcoholic sales, number of alcoholic units, and then, unit size by sales size category, for 7 categories.

Three of the 11 categories are special names. They are primarily takeout, snack stand, and social caterer. I will read the full list of categories and the percent of sales concentration in each one.

After the success of the concept, the category analysis was extended to market segments beyond SIC 5812, so I begin to get total market size by these 11 groups. Now, the beauty of the approach is that anyone can define any set of restaurants using the criteria and make the definition suit their needs. For example, you could define restaurants just on the basis of price. This would be useful for someone like American Express for their credit cards. Needless to say, this classification effort would not be possible without the detailed data from the economic census, particularly miscellaneous subjects. I use the data as control points and use other industry data to allocate within control points. This particular piece of work reflects the motto of my firm "common sense quantified."

Restaurant Business Spectrum, Including 11 Restaurant Categories

Food service sales	Total	Primarily take-out (KFC)	Snack stand (Orange Julius)	Limited menu: Low price self service (McDonald's)	Limited menu: Low moderate price self service (Bonanza)	Limited menu: Moderate price service (Steak & Ale)	Full menu: Low moderate price service (Sambo's)	Full menu: Low moderate price self service (Morrison's)	Full menu: Moderate price service (Valle's)	Full menu: Moderate high price service (Brennan's)	Luxury menu high price continental service ("21" Club)	Social caterer
Total sales	41,609,984	3,956,756	832,086	7,724,127	1,310,714	3,698,379	12,805,163	2,339,932	4,529,967	2,826,584	627,419	958,857
Percent of sales	100.00	9.51	2.00	18.56	3.15	8.89	30.78	5.62	10.89	6.79	1.51	2.30
Number of units	208,668	23,416	7,787	49,867	3,130	24,531	71,232	8,222	9,110	6,588	841	3,944
Percent of units	100.00	11.22	3.73	23.90	1.50	11.76	34.14	3.94	4.36	3.16	.40	1.89
Type												
Total food sales	36,163,013	3,739,134	763,480	7,136,596	1,207,931	3,354,689	11,220,577	2,192,383	3,381,654	1,964,385	337,199	864,985
Percent of sales	100.00	10.34	2.11	19.73	3.34	9.28	31.04	6.06	9.35	5.43	.93	2.39
Take-out food sales	6,821,689	3,549,863	307,615	1,751,856	93,021	355,398	396,761	105,355	(X)	(X)	(X)	262,420
Percent of sales	100.00	52.04	4.51	25.68	1.36	5.21	5.81	1.54	(X)	(X)	(X)	3.85
Take-out food units	102,292	23,416	7,787	31,014	973	11,073	23,071	2,667	(X)	(X)	(X)	2,291
Percent of units	100.00	22.89	7.61	30.33	.95	10.82	22.55	2.61	(X)	(X)	(X)	2.24
Alcoholic sales	3,851,129	(X)	28,981	317,645	26,027	226,186	1,185,856	51,612	1,008,415	754,958	186,343	65,106
Percent of sales	100.00	(X)	.75	8.25	.68	5.87	30.80	1.34	26.18	19.60	4.84	1.69
Alcoholic units	51,269	(X)	1,191	6,502	468	4,625	21,329	964	8,659	6,274	831	426
Percent of units	100.00	(X)	2.32	12.68	.91	9.02	41.61	1.88	16.89	12.24	1.62	.83
Volume												
Less than $49,999	2,155,681	357,848	68,442	380,590	(X)	288,849	967,849	63,937	(X)	(X)	(X)	28,166
Number of units	58,753	8,416	2,317	15,334	(X)	8,209	22,201	1,581	(X)	(X)	(X)	695
$50,000 to $99,999	3,665,086	556,857	112,470	761,866	9,472	447,367	1,455,631	154,512	23,670	20,625	(X)	122,616
Number of units	46,512	6,315	1,750	11,934	121	6,095	16,435	1,709	306	275	(X)	1,572
$100,000 to $299,999	14,464,011	1,218,483	514,303	2,807,734	226,578	1,175,943	6,512,414	728,324	675,195	326,326	23,736	254,975
Number of units	69,536	5,685	3,322	15,973	1,192	7,056	26,596	2,942	3,456	2,119	110	1,085
$300,000 to $499,999	7,471,588	765,678	105,245	1,312,636	439,837	626,725	1,710,977	467,941	1,154,308	728,449	61,780	98,012
Number of units	18,752	1,687	347	3,597	1,053	1,735	3,817	1,075	2,842	2,207	140	252
$500,000 to $999,999	8,870,435	947,003	31,626	1,625,179	458,261	853,307	1,601,093	623,350	1,437,063	870,667	299,643	123,243
Number of units	12,307	1,246	51	2,479	647	1,252	1,927	756	1,898	1,434	441	176
$1,000,000 to $1,999,999	3,760,499	91,801	(X)	668,327	176,566	266,053	366,520	260,024	885,563	675,452	200,928	169,265
Number of units	2,451	60	(X)	494	117	172	210	147	511	487	138	115
$2,000,000 or more	1,222,684	19,086	(X)	167,795	(X)	40,135	190,679	41,844	354,168	205,065	41,332	162,580
Number of units	357	7	(X)	56	(X)	12	46	12	97	66	12	49
Miscellaneous establishments												
Bars and taverns	5,923,911	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)
Number of units	44,112	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)
Hotels, motels, and motor hotels:												
Less than 100 rooms	849,336	(X)	(X)	(X)	(X)	(X)	690,456	(X)	31,480	(X)	(X)	127,400
Number of units	12,767	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)
100 to 299 rooms	1,625,057	(X)	(X)	(X)	(X)	114,754	974,035	(X)	81,252	(X)	(X)	455,016
Number of units	12,933	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)
300 or more rooms	1,526,437	(X)	(X)	(X)	(X)	91,582	500,677	(X)	186,225	120,588	16,790	610,575
Number of units	3,787	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)
Department stores	725,642	(X)	174,154	(X)	(X)	(X)	333,797	203,179	14,512	(X)	(X)	(X)
Number of units	3,882	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)
Variety and general merchandise	556,483	(X)	83,473	(X)	(X)	(X)	473,010	(X)	(X)	(X)	(X)	(X)
Drug and proprietary stores	458,315	(X)	(X)	(X)	(X)	(X)	458,315	(X)	(X)	(X)	(X)	(X)
Other specialized retail stores	737,491	276,611	303,112	57,161	(X)	(X)	100,670	(X)	(X)	(X)	(X)	(X)
Places for special events	918,027	670,159	139,345	(X)	(X)	(X)	62,622	(X)	27,541	(X)	(X)	(X)
Drive-in movie theaters	98,634	98,634	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)
Bowling lanes	294,869	(X)	120,896	(X)	(X)	(X)	173,973	(X)	(X)	(X)	(X)	(X)
Sporting and recreational camps	78,892	(X)	(X)	(X)	(X)	(X)	78,892	(X)	(X)	(X)	(X)	(X)

Note: Restaurant name in parenthesis is paradigm for that category.

(X) Not applicable.

*Unit count includes only those establishments serving food; however, sales figure is for all bars and taverns with payroll.

Other data sources in the food service industry which are representative of available information are:

USDA study of 1969 on types and quantities of food used by 14 market segments. This study is the only comprehensive national study of food consumption by market segment by product. It is subject to serious errors in the size of the different segments. However, happily for us, the key point for the study is the validity of the distribution of the food items within a market segment. On this score, the data seem to be relatively hard. The size of the different market segments and the volume of the products passing through them can be successfully adjusted up or down using the economic censuses as control points: I have used these data in making current estimates of the market size for specific products and have usually come within 10 percent of current surveys.

Nation's Restaurant News features an excellent chain analysis in its two August issues.

Restaurant Business magazine features a projection of market segments of the industry 5 years out (which I prepared) and a most valuable, local-area, detailed list of restaurant and fast-food data generated by Market Statistics, Inc. These data appear in the September issue.

The most comprehensive work on consumer behavior on an on-going basis is provided by CREST, a rotating panel of 10,000 households. Data is obtained by subscription. Some of the data is published in the National Restaurant Association, *NRA News*. The data are very detailed, but there are problems due to the panel composition which is light on young singles and higher income households. As these are important, heavy users of restaurants, the data are questionable sometimes.

The current retail trade program is undergoing a shakeout in the revised data and is still adjusting to the new improved sample. I view the new sample as an improvement over the old one, but I am getting some results which show restaurants and lunchrooms as having higher sales gains than fast-food places. This is not consistent with other industry data.

The Census of Retail Trade—A Useful Marketing Tool

John T. Snow
Sears, Roebuck, and Company

My comments this morning will be confined to the census of retail trade—one of the several segments comprising the U.S. economic census. I will present several applications of retail trade data, that I, as a businessman and user of government statistics, believe makes the census of retail trade a useful marketing tool.

MAJOR USES OF RETAIL TRADE DATA

One of the most useful applications of retail trade data is in the analysis and measurement of sales penetration by type of business, product line, and geographic area. In other words, if I am a hardware retailer, what is my share of hardware store business, what is my share of all lawn and garden equipment sold by all retailers, and what is my total market share in San Diego, Chicago, Des Moines, or Buffalo?

Retailers, and many other industries, have traditionally measured their sales performance as a percentage increase, or decrease, over the previous month or more commonly, the same month of the previous year. Today the sales results of the nation's leading retailers all reported a certain percentage increase over last year. But what does this really mean? Is it up 15 percent from the best month of sales the company had or is it up 15 percent from the depths of a sales slump.

In measuring individual market area performance, retailers and others have fallen into the same trap. The store or market area manager who records a 15 percent increase receives applause while the manager with a 5 percent gain gets little positive attention. But let's look at this on the basis of sales performance relative to market performance. Conceivably, the manager recording a 15 percent sales improvement did so in the midst of a 15 percent, or perhaps a 20 percent, total market growth. In the first instance, he has maintained a market share and in the second he has actually suffered a market share decline. At the same time, the manager with only a 5 percent sales increase may have done so in a stable or even declining market, and, therefore, he actually maintained or increased his market share. With this added dimension—who is the hero and who is the goat?

A continuous policy of beating last year may provide a period of success for a retailer, or other type of business, but it doesn't tell the whole story. In the presence of a steadily rising market, merely beating last year may not be enough and could actually result in an erosion of market share.

Measuring sales performance only on a national basis is also inadequate. Perhaps sales increases in excess of market growth are only occurring in geographic areas where the company faces little direct competition, while in the major markets where competition is more intense, sales performance is falling below market growth and market share is declining—a problem needing immediate corporate attention.

For many years, Sears has effectively used the census of retail trade data to develop a dollar volume by individual geographic market for the total merchandise mix sold by Sears, that is, the type of products that we sell with appropriate product class breakdowns in the durable and nondurable goods categories. Computing our sales in these categories against the total dollar amount sold in that the market allows us to develop a share of market figure that we can trend over time and which we can compare from one geographic market to another of similar size and similar competitive concentration.

Yearly, we compute these market share statistics using the census of retail trade data as benchmarks every 5 years and estimating the intervening years from various government sales tax figures and data purchased from private statistical sources.

To tailor the total retail sales volume, available from the census, to the volume of just those products that Sears sells, we utilize the product line statistics from the census and apply carefully derived weights and ratios to arrive at the desired product mix. As an example, we need the paint, hardware, tool, and some miscellaneous product volume in the building material dealer category but we need to exclude some products such as dimension lumber which we do not sell. The methods we have developed to accomplish this have proven very successful over a relatively long period of time and we have confidence in the results we are obtaining. Incidentally, this process involves literally hundreds of thousands of individual calculations. We have had the process entirely computerized for several years.

Time will not permit me to delve into some of the other major uses of retail trade census data to the extent I have in this instance, but I will at least mention the most important uses that I see.

Determining market potential is directly related to the use I have just described. The marketeer is able to build the volume of his or her type of product or product mix by geographic market and can then determine the desirability of locating a sales facility or marketing effort in that market or abandoning it for another market where the grass appears "greener." I don't mean to suggest that company decisions, to enter a specific market or not, should be based solely on census of retail trade data. That would be sheer folly. However, it can be a useful screening tool enabling the marketeer to measure and rank a large number of markets and then select the most attractive ones for further analysis.

Another major use of the census of retail trade data is in economic and sales forecasting. The retail segment of the nation's business activity is vitally important and requires constant monitoring. National and individual market data from the census can provide a historical base to aid in making both long and short range forecasts. Government data on retail sales collected weekly and monthly on a sample basis are necessary for current information, but the census of retail trade can help in building the historic base.

The planning process, both short range tactical and long range strategic can benefit from the retail trade census data, again, as a historical base, as a market measurement tool, and in trending the changes in particular lines of trade over the long term both nationally and by market.

Related to the planning process, the development of strategies for product or business diversification can also benefit from the retail trade data in measuring markets or businesses that are under consideration.

In somewhat the same vein, the census of retail trade data can be useful in the location of retail stores. There are numerous market conditions and demographic factors that require careful evaluation in making the correct store location decision. The census of retail trade data can provide good market screening and data base, and while not the major input for a store location decision—it should certainly be a part of the statistical evaluation.

Finally, we have on certain occasions been able to use the census of retail trade data to help in a sample selection for marketing reasearch studies. In this regard, this has helped select particular markets and types of business outlets.

SHORTCOMINGS OF THE CENSUS OF RETAIL TRADE

I have tried to outline a few of the major uses of the data from the census of retail trade—applications properly used, can make the census a useful marketing tool. However, I do not want to leave you with the impression that the census of retail trade, in its present format, is without shortcomings, some of which greatly hamper its usefulness. Where can improvements be made? I think that the two most important areas for improvement are in the *timeliness* of the reports and in the *classifications* included. First, the issue of *timeliness*. In 1973, data was collected on the business conducted during 1972. In January 1976, we still did not have some of the needed reports containing 1972 data. Considering the computational, printing, and distribution workload coming up with the 1980 Decennial Census, I will not be surprised if the final printed material from the 1977 Census of Retail Trade is even longer in reaching users. As a user of the retail trade census, I would prefer to have data collected more often than every 5 years. I know that this is an unreasonable expectation; however, I would hope that in the very near future, we could have a much accelerated release of data after the collection period.

Next, the problem of *classification* of data. The Census Bureau has been slow in recognizing new trends in retailing and adjusting the census of retail trade business classifications. By 1963, a census year, discount stores were growing rapidly and it was quite certain that they would remain a viable retailing entity. By 1967, the next census year, they had definitely become a significant force on the retail scene and by 1972, the giants of discounting were well on their way to eclipsing many of the conventional retailing chains. However, we still had no breakout of the discount store segment. I am delighted with the news that the 1977 census will finally have a discount store classification.

A classification problem also exists with food stores. For decades, the supermarket has dominated food retailing, and yet, it is still impossible to distinguish the supermarket from the remaining "Mom and Pop" stores or the new convenience stores.

Hardware chains such as True Value, Pro, Sentry, and Ace have become the major factors in the hardware store industry—again no separate store classification. I am aware of the need to retain comparability of data and classifications from one census to the next, but the loss of classification of such major retailing factors as discount stores and supermarkets seems to be a greater loss to users than the lack of exact comparability to earlier censuses.

Another classification problem rests with the ability of individual businesspersons to self-classify their business. As an example, we don't know how many variety stores are actually classified as such and how many are reported by their managers/proprietors as other kinds of business.

I know it is easy to criticize and I have offered the foregoing shortcomings in a constructive sense. The Census Bureau is cognizant of these problems. I am certain that they are striving to find workable solutions.

CENSUS TRACT PAPERS

Series GE-40

No. 1. Papers Presented at the Census Tract Conference, September 10, 1965 Philadelphia, Pa.
No. 2. Needs and Plans for the 1970 Censuses (Census Tract Conference of August 15, 1966, at Los Angeles, Calif.)
No. 3. Some Uses of Census Tracts in Private Business, by Wilbur McCann (1967)
No. 4. Papers Presented at the Conference on Small-Area Statistics, American Statistical Association, Washington, D.C., December 27, 1967, and Related Papers
No. 5. Papers Presented at the Conference on Small-Area Statistics, American Statistical Association, Pittsburgh, Pa., August 23, 1968, and Related Papers
No. 6. Final 1970 Census Plans and Four Programing Systems for Computerized Data Retrieval and Manipulation (Conference on Small-Area Statistics, American Statistical Association, August 21, 1969, New York, N.Y.)
No. 7. New Uses of Census Resources, The Southern California Census Use Study and a Related Paper (Conference on Small-Area Statistics, American Statistical Association, December 29, 1970, Detroit, Mich.)
No. 8. Small-Area Statistics: Strengthening Their Role in Federal Government and Their Use in Criminal Justice Programs (Papers Presented at the Conference on Small-Area Statistics, American Statistical Association, August 23, 1971, Ft. Collins, Colo.)
No. 9. Social Indicators for Small Areas (Papers Presented at the Conference on Small-Area Statistics, American Statistical Association, August 14, 1972, Montreal, Canada)
No. 10. Statistical Methodology of Revenue Sharing and Related Estimate Studies (Papers Presented at the Conference on Small-Area Statistics, American Statistical Association, December 27, 1973, New York, N.Y.)

Copies of early editions are no longer in print but may be available in Federal Depository Libraries. Papers presented at conferences held from 1958 through 1964 were originally published by the Bureau of the Census in its Working Paper Series.

SMALL-AREA STATISTICS PAPERS

Series GE-41

No. 1. Intercensal Estimates for Small Areas and Public Data Files for Research (Papers Presented at the Conference on Small-Area Statistics, American Statistical Association, August 26, 1974, St. Louis, Mo.)
No. 2. Business Uses of Small-Area Statistics and Education's Needs and Methods for Estimating Low-Income Population (Papers Presented at the Conference on Small-Area Statistics, American Statistical Association, August 25, 1975, Atlanta, Ga.)
No. 3. Statistical Issues in Allocating Federal Funds and Estimation of Local Government Finances (Papers Presented at the Conference on Small-Area Statistics, American Statistical Association, August 25-26, 1976, Boston, Mass.)
No. 4. Interrelationship Among Estimates, Surveys, and Forecasts Produced by Federal Agencies (Papers Presented at the Conference on Small-Area Statistics, American Statistical Association, August 17, 1977, Chicago, Ill.)
No. 5. Methodology and Use of Small-Area Statistics in Decisionmaking and 1977 Economic Censuses and Their Use in Private and Public Sectors (Papers Presented at the Conference on Small-Area Statistics, American Statistical Association, August 15, 1978, San Diego, Calif.)

For Information on any of these reports, write to:

Subscriber Services Section (Publications)
Bureau of the Census
Washington, D.C. 20233

Superintendent of Documents
U.S. Government Printing Office
Washington, D.C. 20402

Official Business

Postage and Fees Paid
U.S. Department
of Commerce
Special Fourth-Class
Rate—Book
COM-202